Evolutionary and population genomic analyses of the *Zymoseptoria* species complex

Dissertation

in the fulfilment of the requirements for the degree “Dr. rer. nat.” of the Faculty of Mathematics and Natural Sciences at the Christian Albrechts University of Kiel

submitted by

Christoph Jakob Eschenbrenner

Kiel, May 2019

Title Picture: Multiple sequence alignment of random sequences generated by CJE as an example for genomics data.

First referee:	**Prof. Dr. Eva Holtgrewe Stukenbrock**
Second referee:	**Prof. Dr. Tal Dagan**
Date of the oral examination:	**24.06.2019**

Zusammenfassung

Seit ihrer Entwicklung vor circa 11,000 Jahren hat sich die Landwirtschaft zum wichtigsten Faktor der menschlichen Nahrungsindustrie entwickelt. Nahrungssicherheit ist somit eines der wichtigsten Themen unserer Gesellschaft. Dabei stellen pflanzenpathogene Pilze eine große Gefahr für diese Nahrungssicherheit dar, in dem sie weltweit zum Teil verheerende Ernteausfälle verursachen. Demzufolge ist die Neu- und Weiterentwicklung von geeigneten Management Strategien ein wichtiges Gebiet der landwirtschaftlichen Forschung. Da landwirtschaftliche Nutzflächen im Gegensatz zu natürlichen Umgebungen einen sehr diversen und hoch verwalteten Lebensraum darstellen, ist es für die Entwicklung eines effektiven Krankheitsmanagements von gesteigerter Wichtigkeit die Unterschiede zu untersuchen, die zwischen erfolgreich angepassten pflanzenpathogene Pilze in beiden Umgebungen bestehen. Eine der Hypothesen dieser Arbeit ist, dass diese Unterschiede zum Beispiel demographische Effekte, wie alternierende Populationsgrößen und Populationsstrukturen, Veränderungen der Genomstruktur oder funktionelle Anpassungen bestimmter Gene, welche die Fitness im entsprechenden Lebensraum erhöhen, umfassen. Ziel dieser Arbeit ist es, die oben beschriebenen Unterschiede zwischen verwalteten und natürlichen Ökosystemen innerhalb des *Zymoseptoria* Artkomplexes zu untersuchen. Dabei wurden sowohl Methoden der Molekulargenetik, als auch der Populationsgenomik und der Vergleichenden Genomik eingesetzt.

Die Ergebnisse der Untersuchungen zeigen, dass alle Vertreter des *Zymoseptoria* Artkomplexes eine ungewöhnliche Genomplastizität aufweisen, die sich durch das gattungsweite Auftreten akzessorischer Chromosomen und durch strukturelle Veränderungen die mit repetititven Elementen assoziiert sind manifestiert. Anhand von Untersuchungen der genetischen Diversität konnten ebenfalls starke Unterschiede zwischen den unterschiedlichen Ökosystemen gefunden werden. So zeigt sich, dass der Weizenpathogen *Z. tritici* eine sehr viel größere effektive Populationsgröße (genetische diversität) entwickelt, als die nah verwandten Schwester-Spezies *Z. ardabiliae* und *Z. brevis*. Unter Verwendung populationsgenetischer Methoden bestätigt diese Dissertation, dass die lokalen Populationen des Weizenpathogens *Z. tritici* Teil einer globalen Großpopulation in Panmixie und einer schwachen Populationsstruktur sind. Ferner zeigen die Ergebnisse, dass die effektive globale Populationsgröße von *Z. tritici* seit der Artbildung stark angestiegen ist. Im Gegensatz dazu weisen die Datensätze der Wildgras-assoziierten Pathogene auf Unterschiede in den Populationsstrukturen,

niedrigere genetische Diversität und daraus ableitend Veränderungen der Entwicklung der effektiven Populationsgröße auf.

Die genomweite Suche nach Genen mit einer potentiellen Funktion bei der Anpassung an die Umwelt resultierte in zwei Listen aus Genen mit Signaturen innerartlicher diversifizierender und zwischenartlicher direktionaler positiver Selektion sowie in einen Datensatz Effektor Kandidaten. Die Vergleiche der Listen zeigen dabei, dass positive diversifizierende und direktionale Selektion im landwirtschaftlichen Ökosystem eine größere Rolle zu spielen scheinen als in den natürlichen. Darüber wird ein Gen mit Signaturen beider Arten positiver Selektion, ein potentieller ABC-Transporter, identifiziert. Ferner deutet diese Studie auf die Wichtigkeit artspezifischer Effektoren für die Anpassung an das Ökosystem hin. Weitergehende experimentelle Analysen wie Gendeletionen gefolgt von *in planta* Analysen sind nötig, um die Bedeutung der Effektor-Kandidaten und der Gene mit Signaturen positiver Selektion für die Virulenz und die Umweltanpassung der *Zymoseptoria*-Arten näher zu untersuchen.

Abschließend ist hervorzuheben, dass die großen Unterschiede in der genetischen Diversität zwischen den *Zymoseptoria*-Arten aus natürlichen und landwirtschaftlichen Ökosytemen in weitergehenden Untersuchungen der ökologischen Unterschiede dieser Umgebungen dazu genutzt werden könnten, um die maßgeblichen Determinanten zu bestimmen, welche die effektive Populationsgröße des Pathogens beeinflussen. Dadurch könnten Kerneigenschaften der natürlichen Umgebungen wie zum Beispiel eine höhere genetische Diversität des Wirts in zukünftige Strategien gegen *Z. tritici* integriert werden.

Summary

Since its development about 11,000 years ago, agriculture has become an essential factor in the human food industry. Food security is thus one of the most important topics of our society. However, phytopathogenic fungi represent a significant threat to food security, causing devastating crop failures worldwide. As a result, the new and further development of appropriate management strategies is an essential field of agricultural research. Since agricultural land, in contrast to natural environments is a very diverse and highly managed habitat, it is of increased importance for the development of effective disease management to investigate the differences that exist between successfully adapted phytopathogenic fungi in both environments. One of the hypotheses of this work is that these differences include, for example, demographic effects, such as alternating population sizes and population structures, changes in the genome structure, or functional adaptations of particular genes that increase fitness in the corresponding habitat. This work aims to investigate the, above mentioned, differences between managed and natural ecosystems within the *Zymoseptoria* species complex. For this, it used both methods of molecular genetics, as well as population genomics and comparative genomics.

The results of the investigations show that all members of the *Zymoseptoria* species complex show unusual genome complicity, which is manifested by the genus-wide occurrence of accessory chromosomes and by structural changes associated with repetitive elements. Genetic diversity studies also revealed substantial differences between the different ecosystems. Thus, the wheat pathogen *Z. tritici* develops a much larger effective population size (genetic diversity) than the closely related sister species *Z. ardabiliae* and *Z. brevis*. Using population genetic methods, this dissertation confirmed that the local populations of the wheat pathogen *Z. tritici* are part of a large global population in panmixia and a weak population structure. Further, the results show that the effective global population size of *Z. tritici* has dramatically increased since speciation. In contrast, records of wild grass associated pathogens show differences in population structures, lower genetic diversity and, as a consequence, changes in the evolution of effective population size.

The genome-wide search for genes with a potential function in adapting to the environment resulted in two lists of genes with signatures of intra-species diversifying and inter-species directional positive selection as well as a data set of effector candidates. The comparisons of the lists show that positive diversifying and directional selection

seems to play a more prominent role in the agricultural ecosystem than in the natural ones. Also, a gene with signatures of both types of positive selection, a potential ABC transporter, is identified. Further, this study suggests the importance of species-specific effectors for adaptation to the ecosystem. Further experimental analyzes such as gene deletions followed by *in planta* analyzes are needed to further investigate the importance of effector candidates and genes with signatures of positive selection for virulence and environmental adaptation of the *Zymoseptoria* species.

Finally, it should be emphasized that the substantial differences in genetic diversity between *Zymoseptoria* species from natural and agricultural ecosystems could be exploited in further studies of the ecological differences of these environments to determine the critical determinants that influence the effective population size of the pathogen. This would allow core properties of natural environments, such as a higher genetic diversity of the host, to be integrated into future strategies against *Z. tritici*.

Table of Contents

List of abbreviations

#	Number	
%	Percent	
°C	Degree Celsius	
aa	Amino acid	
ABC-transporter	ATP-binding cassette transporters	
AMOVA	Analysis of molecular sequencing	
avr	avirulence gene	
Avr-R	Avirulence gene - resistance gene recognition	
BAM	Binary sequence alignment/map format	(Data format)
BEB	Bayes empirical Bayes	
BED	Browser extensible data	(Data format)
bp	Basepair	
BWA	Burrows-Wheeler Aligner	
CDS	Coding sequence	
cm	Centimeter	
CSV	Comma-separated value	(Data format)
dN	Number of non-synonymous substitutions	
Dn	Number of non-synonymous substitutions	
DNA	Deoxyribonucleic acid	
dS	Number of synonymous substitutions	
Ds	Number of synonymous polymorphisms	
EDTA	Ethylenediaminetetraacetic acid	
ETI	Effector-triggered immunity	
ETS	Effector-triggered susceptibility	
FDR	False discovery rate	
GATK	Genome analysis toolkit	
Gb	Gigabyte	
GBS	Genotyping by sequencing	
GFF	Gene feature format	(Data format)
h	Hour	
H20	Water	
HMM	Hidden Markov model	
HR	Hypersensitive response	
i.e.	For example	
ID	Identifier	
ISM	Infinite sites model	
Kb	Kilobase	
LD	Linkage disequilibrium	
LRT	Likelihood ratio test	
LS	Lineage-specific regions	
MAF	Multiple alignment format	(Data format)
MAMP	Microbe associated molecular pattern	

Mb	Megabase	
MGA	Multiple genome alignment	
min	Minute	
MK test	McDonald-Kreitman test	
mL	Milliliter	
MSMC	Multiple sequentially markovian coalescent	
Nc	Census population size	
Ne	Effective population size	
NEB	Naive empirical Bayes	
NI	Neutrality index	
PAMP	Pathogen associated molecular pattern	
PCA	Principal component analysis	
PCR	Polymerase chain reaction	
PED	Plink data format	(Data format)
PFGE	Pulsed-field gel electrophoresis	
Ph.D.	Doctor of philosophy	
Pn	Number of non-synonymous polymorphisms	
PRR	Pattern recognition receptors	
Ps	Number of synonymous polymorphisms	
PTI	PAMP-triggered immunity	
R	Resistance Gene	
RADseq	Restrictive associated sequencing	
rpm	Rounds per minute	
S	Number of segregating sites	
s	Seconds	
SAM	Sequence alignment/map format	(Data format)
SDS	Sodium dodecyl sulfate	
SFS	Site frequency spectrum	
SMRT	Single molecular real-time sequencing	
SNP	Single nucleotide polymorphism	
TBE	Tris/Borat/EDTA	
TE	Transposable element	
Tris-HCl	TRIS hydrochloride	
USB	Universal serial bus	
V	Volt	
VCF	Variant calling format	(Data format)
WGS	Whole genome sequencing	
YMS	Yeast malt sucrose buffer	
θ	Watterson estimator of nucleotide diversity	
μ	Mutation rate	
μg	Microgram	
$\boldsymbol{\pi}$	Nucleotide diversity	

1. Introduction

1.1. Natural and managed ecosystems: contrasting ecological conditions

Crop domestication and agricultural practices were developed about 11,000 years ago in the Fertile Crescent a region in the middle east (Zeder 2011). Since then, the agriculturally used area increased to up to 38% of earth terrestrial surface (Foley et al. 2011). This expansion was possible due to the switch from monocultures of locally adapted but genetically diverse landraces towards monocultures of highly adapted and genetically uniform crop varieties (Evenson and Gollin 2003; Van De Wouw et al. 2010; Pingali 2012). As a result of these developments, agricultural ecosystems are characterized by high host densities with reduced genetic diversity, while natural ecosystems show the opposite (Burdon 1993; Stoate et al. 2001; Gilbert 2002; Burdon and Chilvers 2003; Gurr et al. 2003; Burdon and Thrall 2008; Moonen and Bàrberi 2008; Stukenbrock and McDonald 2008). Because of these differences, agricultural ecosystems place entirely different selection regimes on plant pathogens than natural habitats (McDonald and Stukenbrock 2016). In modern agricultural ecosystems, plant pathogens are regularly confronted with pesticides, changing host resistances and crop rotations (Stahl and Bishop 2000; McDonald and Linde 2002; Zhan et al. 2002). It has already shown that the higher host density in agricultural ecosystems could lead to increased virulence in phytopathogenic fungi through increased efficacy in transmission between hosts (Read 1994) or the possible occurrence of multiple infections (Keller et al. 1997; McDonald et al. 1999; Linde et al. 2002) and co-infections (Alizon et al. 2013).

Furthermore, global trade offers the possibility of the emergence and spread of new pathogens (Goodwin et al. 1994; Goss et al. 2011). Species of closely related phytopathogenic fungi adapted to natural or agricultural ecosystems, such as species of the genus *Zymoseptoria*, provide an ideal model system for the analysis of speciation and host adaptation (Stukenbrock et al. 2012). By using methods of comparative genomics and population genomics Stukenbrock and co-workers could show recent divergence times of up to 22,000 years in this genus (Stukenbrock et al. 2007; Stukenbrock et al. 2012).

A detailed description of the genus *Zymoseptoria* will be given elsewhere in this introduction.

Using methods of comparative genomics and population genomics this work aims to examine to what extent the adaptation of phytopathogenic fungi to natural or agricultural ecosystems influences population structures, demographic trends and the evolution of individual genes.

Below, this introduction gives an overview of the biology of fungal plant pathogens, followed by a presentation of the genus *Zymoseptoria* (1.2). Furthermore, the methods and principles of comparative genomics and population genomics that were used here to answer the question are highlighted (0). Finally, a brief overview of the preparation of a suitable data set for population genomic analyzes is given (1.4).

1.2. Plant pathogenic fungi – A threat to plant health and agriculture

Epidemic plant diseases caused by fungi and oomycetes represent global threats to food security. Extensive farming practices, homogenization of landscapes by agriculture, global trade and transportation promote this effect (Fisher et al. 2012). In 2011 simulations of disease scenarios indicate that fungal plant pathogens can cause gain losses that could feed up to 8.5% of the global human population (Fisher et al. 2012). Bacterial and fungal pathogens cause up to 16% grain losses in crop production (Oerke 2006). Both ascomycetous and basidiomycetous fungi are amongst the top 10 fungal plant pathogens threatening agricultural food production (Dean et al. 2012). As a consequence of speciation and host specialization, fungal plant pathogens developed large varieties of host ranges, lifestyles, genome sizes and genome composition (Table 1).

Where and how to live? Host ranges and lifestyles

Host ranges of fungal plant pathogens can be classified as narrow or broad, defined by the number of host species that the can successfully colonize (Schulze-Lefert and Panstruga 2011).

A fungal plant pathogen with broad host range is, for example, the causing agent of the grey mold fungus *Botrytis cinerea* that can infect over 200 different plant species (Williamson et al. 2007). On the other hand, many fungal plant pathogens are highly

specialized, such as the wheat pathogen *Zymoseptoria tritici* (syn. *Mycosphaerella graminicola*). Host ranges are also dynamic and can change via host jumps. The ascomycete fungal plant pathogen *Magnaporthe oryzae* causing rice blast on rice was reported in 1985 also to infect wheat (*Triticum aestivum L.*) in Brazil (Igarashi et al. 1986). Recently wheat infecting lineages were reported to causes severe epidemics in Bangladesh (Islam et al. 2016; Malaker et al. 2016).

Depending on their lifestyle fungal plant pathogens can be categorized as biotrophic, hemibiotrophic and necrotrophic (Jones and Dangl 2006). In general, these lifestyles the grade of damage the pathogen causes to the host plant separates these lifestyles from each other. Necrotrophic fungi (Wolpert et al. 2002; Friesen et al. 2006; Jones and Dangl 2006) rapidly kill their host after successful colonization using specialized toxins and enzymes, whereas biotrophic fungi require a living host to complete their life cycle (Jones and Dangl 2006; Lo Presti et al. 2015). Hemi-biotrophic fungi represent an intermediate lifestyle characterized by an initial phase of biotrophy followed by a switch to necrotrophy (Dangl and Jones 2001; Stukenbrock et al. 2007). However, independently of the lifestyle, the pathogen must be able to overcome and escape the immune system of the host.

The host immune system operates on different levels by physical barriers or proteins that target pathogen structures to control the pathogen (Jones and Dangl 2006; Katagiri and Tsuda 2010; Javelle et al. 2011). Plants can successfully fight off most pathogens since the majority of non-specialized pathogens cannot overcome the physical barriers which represent the first layer of plant defenses (Dangl and Jones 2001). Upon recognition of the pathogen by the host plant a complex antagonistic interaction between the host immune system and the pathogen begins. The zig-zag model developed in 2006, describes the interaction between pathogen and host (Jones and Dangl 2006) (Figure 1).

The zig-zag model states that host plants can recognize slowly evolving microbe- or pathogen-associated molecular patterns (MAMPS or PAMPs) such as flagellin (bacteria) or chitin (fungi) by transmembrane pattern recognition receptors (PRRs) (Zipfel and Felix 2005; Jones and Dangl 2006; Wan et al. 2008). Recognition of PAMPs by the host can lead to PAMP-triggered immunity (PTI) and in many cases prevent the successful colonization by the pathogen (Jones and Dangl 2006). Fungal effectors, highly specialized proteins or small molecules can manipulate the recognition of PAMPS or the subsequent signaling by the PRRs (Figure 1). These effectors can

prevent PTI and thereby enable successful infection and confer effector-triggered susceptibility (ETS) of the plant (Jones and Dangl 2006; De Wit 2007). Plant hosts, in turn, can directly or indirectly recognize effectors by NB-LRR (nucleotide binding and leucine-rich) proteins, leading to effector-triggered immunity (ETI), a kind of PTI resulting into a hypersensitive response (HR) causing local cell death.

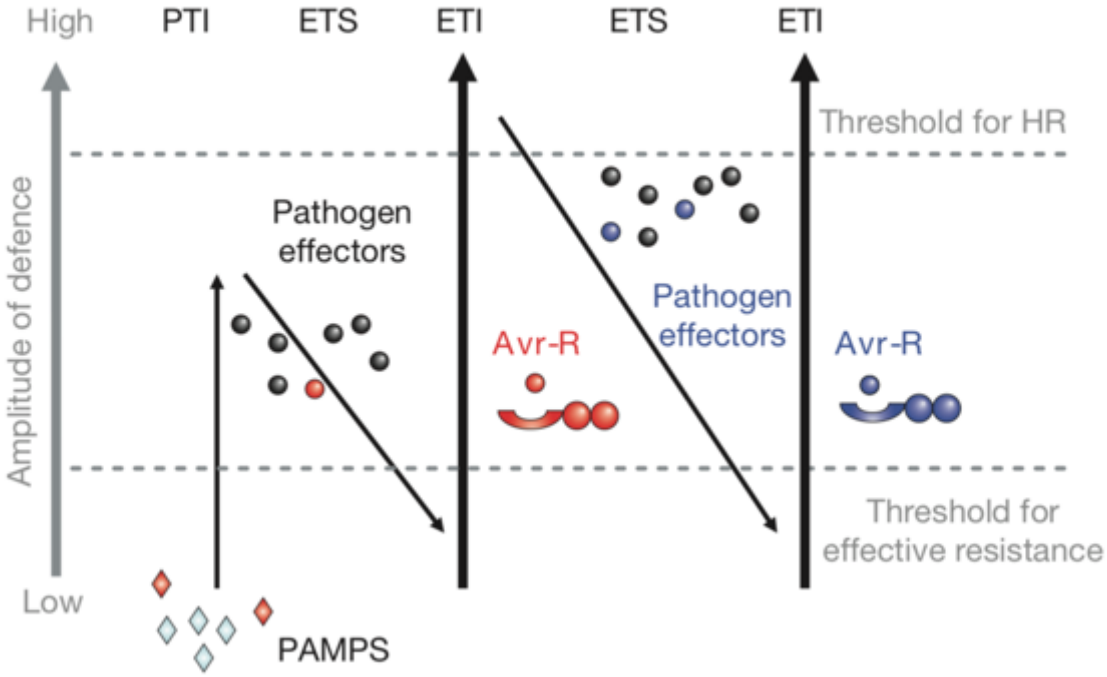


Figure 1 - The zig-zag model of plant immunity against pathogens as proposed by Jones and Dangl (Jones and Dangl 2006). PAMPs (pathogen-associated molecular patterns). PTI (PAMP-triggered-immunity). ETS (Effector-triggered-susceptibility). ETI (Effector-triggered-immunity). Avr-R (recognition of an effector by a resistance gene (R) of the host. The effector is now an avirulence (avr) gene). In the model, the recognition of PAMPs (red diamonds) leads to a defensive reaction in the host (PTI). Subsequently, specialized pathogen effectors break the PTI to overcome the host immune defense. The recognition of a fungal effector (red point) by a resistance gene (R) restores host immunity (Avr-R) and led to ETI. Subsequently, the ongoing evolution of pathogen effectors has the potential to restore ETS, which the host again may end by developing new R-genes.

As an essential characteristic of a constant co-evolution between pathogen and host, natural selection could affect the recognized effectors and thus restore the ETS (Jones and Dangl 2006; Lo Presti et al. 2015). Co-evolution between pathogen and host can follow two alternative dynamics called “trench warfare” and “arms race”. The “trench warfare” dynamic implies a continuously variable pool of alleles in host and pathogen that changes the proportion of distinct alleles periodically over time.

In contrast, the accumulation of advantageous alleles in host and pathogen over time characterizes the “arms race” dynamic (Tellier et al. 2014). Agricultural ecosystems,

however, evolve due to constant anthropogenic artificial selective pressures. Thus, agricultural ecosystems represent a significant challenge to pathogens through dynamic changes in environmental conditions, such as frequent changes in host genotypes and the use of antimicrobial agents (Thrall 2003; Swift et al. 2004; Alizon et al. 2009; Hovmøller et al. 2011; McDonald 2014; Thrall et al. 2015). Consequently, pathogenic species may find different host resistances in every growing season. Frequent exchange of resistant cultivars and varieties of domesticated plants is considered to lead to accelerated evolution towards host adaptation of fungal plant pathogens in agroecosystems (Möller and Stukenbrock 2017).

Functionally, two types of effectors can be categorized. One group of effectors is involved in suppressing the immune system, whereas another actively causes host death (Lo Presti et al. 2015). The nature of the required effectors depends on the lifestyle of the fungal plant pathogen (Lo Presti et al. 2015). Necrotrophs and hemibiotrophs, for example, have effectors such as PtrToxA (*Pyrenophora tritici-repentis*) or NIS1 (*Colletotrichum orbiculare*), which cause necrosis (Manning 2005; Yoshino et al. 2012; Day et al. 2015). The example of the effector PtrToxA also shows the interaction between effector and the susceptibility gene of the host, in this case, Tsn1, that lead to virulence (Manning 2005; Day et al. 2015). On the other hand, biotrophs and hemibiotrophs require effectors that allow them to remain undetected or to defend themselves against possible countermeasures by targeted manipulation of the host immune system. Effectors such as Avr4 (*Cladosporium fulvum*) or Slp1 (*M. oryzae*) prevent, for example, the detection of the immune system by masking fungal chitin (Kaku et al. 2006; Mentlak et al. 2012; de Wit et al. 2012). Other effects, such as Tom1 in *C. fulvum*, actively combat the action of plant defense products (Okmen et al. 2013).

Reproductive modes of fungal plant pathogens

Fungi can reproduce sexually, asexually or in both ways (Table 1). The following describes the impact of sexual reproduction on genome architecture of fungal pathogens. Currently, two hypotheses exist for the evolutionary advantage of sexual propagation: i) According to the Red Queen Hypothesis (Van Valen 1973), by driving the co-evolutionary arms race between host and pathogen by securing rare genotypes in the host population sexual reproduction can be maintained. Because the pathogen population mainly adapts to the most common host genotypes, natural selection will evolve future host populations from the rarer and therefore less susceptible genotypes (Jayakar 1970; Jaenike 1978; Clark 1979; Hamilton 1980). ii) Sexual recombination also

facilitates the breakup of alleles at neighboring positions (Figure 2), also termed Hill-Robertson interference (Hill and Robertson 1966). Consequently, the Hill-Robertson Effect causes a strong linkage between deleterious and beneficial alleles, can break up these ties again (Barton 1995; Otto and Barton 1997; Roze and Barton 2006).

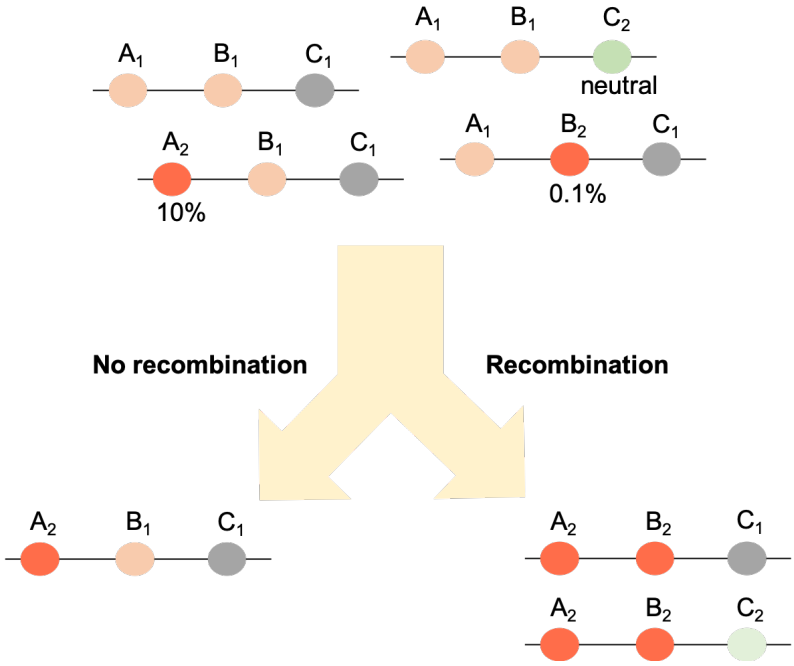


Figure 2 – Recombination breaks up the linkage between neighboring loci. This example considers three linked nucleotide sites A, B, and C. The new advantageous alleles increase fitness by 10% (A2) and 0.1% (B2). Site C is neutral with two neutral alleles C1 and C2. In the absence of recombination (left arrow), A2 gets rapidly fixed in the population because of its considerable fitness advantage. Alleles B1 and C1, on the other hand, get fixed by genetic hitchhiking with the selected A2 allele. In the presence of recombination, more fit combinations (A2B2C2 and A2B2C1) can arise and become predominant in the population. This model shows that the efficacy of selection and neutral diversity are higher with recombination than in its absence. Modified from the original publication (Marais and Charlesworth 2003).

Analyses in *Fusarium graminearum*, *M. oryzae* and *Z. tritici* show high recombination rates and recombination hotspots next to or in coding regions, suggesting that frequent recombination may play a role in the fixation of advantageous mutations (Zheng et al. 2008; Croll et al. 2015; Talas and McDonald 2015; Stukenbrock and Dutheil 2017). In some species like *Ustilago maydis* pathogenicity is directly linked with the ability to finish the sexual cycle, since sterility causes avirulence (Feldbrügge et al. 2004).

High linkage of genetic elements in genomes caused by reduced recombination is a characteristic of asexual/clonal reproduction (reviewed in De Meeûs et al. 2007). Beneficially, clonal reproduction leads to a rapid propagation and fixation of co-adapted genotypes (McDonald and Linde 2002). However, an evolutionary 'downside' of clonality is a possible accumulation of non-beneficial mutations (Muller 1932; Möller and

Stukenbrock 2017). In 1964 Herman Muller hypothesized a process, later known as Muller's ratchet that describes how non-beneficial mutations accumulate in small clonal populations (Muller 1964). He proposed that since in a clonal population of a finite size only new mutations can occur, genetic drift over time may lead to the elimination of genotypes with no or few mutations. This effect leads to the irreversible accumulation of non-beneficial mutations and the fixation of less fit genotypes. Accidental elimination of fit genotypes by the effect of genetic drift is further likely to lower the population size of asexual populations which again, increases the impact of random genetic drift (Gabriel et al. 1993). The effect of asexuality and the resulting low population size in predominantly asexual species may further lead to the accumulation of more repetitive sequenced such as transposable elements (TE), because of reduced recombination (Lynch and Conery 2003; Dolgin and Charlesworth 2006). A higher proportion of transposable elements is a common trait in both sexual and asexual reproducing plant pathogenic fungi (Table 1).

However, what about the evolutionary disadvantages of asexuality? Analyses in *Verticillium dahliae* showed evidence that rare sexual recombination may compensate long periods of asexual reproduction. (Milgroom et al. 2014; Short et al. 2014).

Hybridization (sexual or asexual mating between individuals of different species) and introgression (recurrent backcrossing between hybrid and parent) are considered to be essential processes in the evolution of fungal plant pathogens (Stukenbrock 2016). So far, several studies could provide numerous examples of hybrid fungal plant pathogens that originated from inter-specific mating (Schardl and Craven 2003; Stukenbrock 2016). Population genomic analyses based on a population genomic dataset containing genome-wide SNPs could identify *Blumeria graminis f. sp. triticale* as a hybrid species that originated from hybridization between *B. graminis f. sp. tritici* and *B. graminis f. sp. secalis* (Menardo et al. 2016).

What do fungal plant pathogens offer as models for the evolution of microbial eukaryotes?

Large population sizes, short generation times, small genomes sizes, high morphological and ecological diversity, as well as amenability to experimental approaches, make plant pathogenic fungi ideal model systems for the analyses of evolutionary processes of microbial eukaryotes in a wide range of environmental backgrounds. Additionally, the possibility of comparisons between closely related pathogen species with different

host ranges including and excluding species associated with domesticated plant species can provide insights into the genomic basis of specialization to domesticated environments (Stukenbrock 2013).

Population genomics applies methods of population genetics on genome-wide scales in order to reveal evolutionary and demographic processes that affect population structure or speciation, as well as the divergence of closely related taxa (Grünwald et al. 2016). Also, the analysis of demographic processes may allow conclusions regarding the co-evolutionary dynamics between host and pathogen and potentially support disease management. Comparative genomic approaches between closely related taxa, as well as between isolates of the same species with differing host ranges provide insight into speciation processes and host specialization. These comparative approaches can also identify genomic regions and units that contain potential virulence factors which represent new targets for pathogen control. In conclusion, these population and comparative genomic approaches provide powerful tools for analyzing modes and rates of adaptation. However, plant pathogenic fungi massively challenge current population genomic approaches because of their considerable differences in genome sizes, genome composition, demography and reproductive modes (Table 1).

Genome architecture and composition – Challenges for population genomics

Fungal plant pathogen genomes are comparatively small but are strikingly different in size. Sequenced genomes range from 18 - 180 Mb (Table 1). A large number of sequenced genomes are available for fungal plant pathogens that allow inferences on general aspects of genome organization for these fungi (Stukenbrock and Croll 2014). In particular, an essential driver of the rapid evolution in fungal pathogens is genome plasticity (Fierro and Martín 1999; Raffaele and Kamoun 2012; Galazka and Freitag 2014; Seidl and Thomma 2014). Comparative analyses of 18 fungal plant pathogenic Dothideomycetes revealed a high rate of intra-chromosomal rearrangements characterized by conserved gene content but variation in gene synteny (Hane et al. 2011). This phenomenon has been defined as mesosynteny (Ohm et al. 2012). The mechanisms leading to mesosynteny are so far unknown (Möller and Stukenbrock 2017).

Additionally, repeat content, genes and gene density in the genomes are highly variable (Table 1). Invasions and expansions of TEs are considered to be causal to these broad ranges of TE contents and genome sizes in fungal plant pathogens (Table 1). Accessory chromosomes, also known as supernumerary chromosomes or

conditionally dispensable chromosomes, show inter-specific presence-absence polymorphisms. In many fungal plant pathogens, these accessory chromosomes encode virulence traits and are enriched in repetitive sequences (Stukenbrock et al. 2010; Mehrabi et al. 2017; Möller and Stukenbrock 2017; Soyer et al. 2018). The plant pathogenic fungi *F. solani* and *F. oxysporum* harbor and require lineage-specific chromosomes for pathogenicity on certain host plant species (Ma et al. 2010; van der Does et al. 2016). Evidence of horizontal transfer of these lineage-specific chromosomes between *F. oxysporum* lineages could be exemplary on how asexual organisms could create new combinations of beneficial alleles (Ma et al. 2010).

The genus Zymoseptoria

The ascomycete genus *Zymospetoria* harbors several species adapted to grow and reproduce in the leaf tissue of different graminicolous hosts (Quaedvlieg et al. 2011). A critical member of this genus is the hemibiotrophic pathogen *Z. tritici* causing agent of the Septoria Tritici Blotch (STB) of wheat. *Z. tritici* is a significant thread of wheat production and has been intensively studied (Dean et al. 2012; Fones and Gurr 2015; Torriani et al. 2015). Interestingly, the speciation of *Z. tritici* coincided with the domestication of its host wheat in the region of the Fertile Crescent approximately 12,000 years ago (Banke and Mcdonald 2005; Stukenbrock et al. 2007; Stukenbrock et al. 2011). *Z. tritici* had a narrow host range and evolved a high specialization on wheat by co-evolution (Eyal et al. 1973; Eyal et al. 1985; Van Ginkel and Scharen 1987). During infection, the pathogen enters the host leaf by the stomata and subsequently spreads into the mesophyll (Kema et al. 1996; Rudd et al. 2015; Haueisen et al. 2017). After successful colonization, the fungus can reproduce both sexually by forming pseudothecia or asexually by forming pycnidia (Ponomarenko et al. 2011).

The reference genome of IPO323 (*Z. tritici*) has a total length of 39.7 Mb of a total of 21 chromosomes (Goodwin et al. 2011). Chromosomes 1-13 are categorized as core and 14-21 as accessory chromosomes that show strain-specific presence-absence polymorphisms (Mehrabi et al. 2007; Goodwin et al. 2011). Accessory chromosomes can be commonly found in the genus *Zymoseptoria* (Stukenbrock et al. 2011). Core and accessory chromosomes differ in TE and gene density (Grandaubert et al. 2015),

Table 1 – Genomes of several important plant pathogenic fungi.

Species (Isolate)	Phylum	Natural host	Lifestyle	Reproduction	Genome size (Mb)	Repeats (%)	Genes	References
Blumeria graminis f. sp. hordei (DH14)	Ascomycota	Various grasses	Obligate bio-troph	Sexual/asexual	120	64	5,854	Spanu et al. 2010; Hacquard et al. 2013
Blumeria graminis f. sp. tritici (96224)	Ascomycota	Various	Biotroph	Sexual/asexual	180	90	6,540	Wicker et al. 2013
Colletotrichum fructicola (Nara-gc5)	Ascomycota	Various	Hemi-bio-troph	Asexual	56	0.75	15,469	Crouch et al. 2014
Colletotrichum graminicola (M1.001)	Ascomycota	Maize	Hemi-bio-troph	Asexual	57	12.2	12,006	Crouch et al. 2014
Colletotrichum higginisianum (IMI349063)	Ascomycota	*Brassica Raphanus*	Hemi-bio-troph	Asexual	53	1.2	16,172	Crouch et al. 2014
Colletorichum orbiculare (104-T)	Ascomycota	Curcubits	Hemi-bio-troph	Asexual	88	8.3	13,479	Crouch et al. 2014
Fusarium graminearum (PH-1)	Ascomycota	Wheat Barley	Hemi-bio-troph	Sexual/asexual	36	0.1	11,640	Cuomo, Güldener, Xu et al. 2007
Fusarium oxysporum f. sp. lycopersici (4287)	Ascomycota	Various	Hemi-bio-troph	Asexual	60	28	17,735	Ma et al. 2010
Leptosphaeria maculans 'brassicae'	Ascomycota	Crucifers	Hemi-bio-troph	Sexual/asexual	45	34.2	12,469	Rouxel et al. 2011

Magnaporthe oryzae (70–15)	Ascomycota	Various	Hemi-biotroph	Sexual/asexual	41	10	11,109	Dean et al. 2005; Galazka and Freitag 2014
Puccina graminis f. sp. tritici (CDL75-36-700-3, race SCCL)	Basidiomycota	Wheat Barley	Biotroph	Sexual/asexual	89	45	17,773	Duplessis et al. 2011
Sporisorium reilianum (SRZ2)	Basidiomycota	Maize	Biotroph	Sexual	18	8.3	6,648	Dutheil et al. 2016
Sporisorium scitamineum (Sscl8)	Basidiomycota	Sugercane	Biotroph	Sexual	20	6.7	6,693	Dutheil et al. 2016
Ustilago maydis (521)	Basidiomycota	Maize	Biotroph	Sexual	20	6.7	6,786	Dutheil et al. 2016
Verticilium dahliae (VdLs.17)	Ascomycota	Various	Necrotroph	Asexual	37	1.3	10,535	De Jonge et al. 2013
Zymoseptoria ardabiliae (ST11IR_6.1.1)	Ascomycota	Various grasses	Hemi-biotroph	Sexual/asexual	31.5	7.9	12,719	Grandaubert et al. 2015
Zymoseptoria brevis (Zb87)	Ascomycota	Various grasses	Hemi-biotroph	Sexual/asexual	31.9	23.9	10,649	Grandaubert et al. 2015
Zymoseptoria pseudotritici (ST04IR_5.5)	Ascomycota	Various grasses	Hemi-biotroph	Sexual/asexual	33	15.5	12,027	Grandaubert et al. 2015
Zymoseptoria tritici (IPO323)	Ascomycota	Wheat	Hemi-biotroph	Sexual/asexual	40	18.6	10,933	Goodwin et al. 2011

as well as in histone methylation (Schotanus et al. 2015) and the rate of recombination (Stukenbrock and Dutheil 2018).

Z. tritici has an extremely high standing variation with the genetic diversity in populations of *Z. tritici* to be high on local and global scales (McDonald and Martinez 1991; Linde et al. 2002).

Besides the wheat pathogen *Z. tritici*, the genus *Zymoseptoria* harbors three other closely related sister species *Z. ardabiliae*, *Z. brevis* and *Z. pseudotritici* that infect wild grass species. Interestingly, these sister species were so far only reported from regions in the Fertile Crescent where they initially were described on wild grass species of the plant genera *Dactylis*, *Elymus*, *Phalaris*, *Lolium* and *Agropyron* in the provinces Ardabil and Ilam in Iran (Quaedvlieg et al. 2011; Stukenbrock et al. 2012). Therefore, the genus *Zymospetoria* contains both highly specific co-domesticated plant pathogens with narrow host range and closely related sister species that are infecting several wild grass species. Hence this genus represents a particularly, well-suited model for the analysis of speciation and host specialization of fungal pathogens in agricultural and natural habitats.

1.3. Comparative and Population Genomics

Population genomics

Population genomics can be defined as the study of genome-wide genetic variation in species and populations (Black IV et al. 2002). Neutrally evolving loci across the genome were similarly affected by demography and random genetic drift while loci or genomic regions under selection will behave differently (reviewed in Luikart 2014). Consequently, population genomics provides an excellent tool for the detection of selection at separate loci as well as of evolutionary processes that act on the entire genomes in populations and species (Black IV et al. 2002). Section 1.4 of this thesis gives more information about the generation of a population genomic dataset with a genome-wide variation.

Methods of population genomics use genome-wide genetic variation in species that emerges by mutations. Mutations that produce single nucleotide polymorphisms (SNPs), as well as chromosomal rearrangements, are the source of genetic variation across genomes (Hartl and Clark 2007). They create patterns of genetic variations that can be analyzed using methods of population genomics to understand better how evolutionary processes act on separate loci (selection and recombination) or entire

genomes (random genetic drift, demographic events, gene flow or inbreeding) (Black IV et al. 2002).

One way to explain how patterns of genetic variation are generated and maintained during the evolution of genomes is known as the neutral theory of molecular evolution (Kimura 1968). This pure model states that most polymorphisms in nucleotide composition are generated and maintained by the action of random genetic drift rather than natural selection. Based on the assumptions of the neutral theory, the infinite-sites model (ISM) was developed to explain the emergence of genetic variation in DNA sequences. The model considers a very long nucleotide sequence (infinite sites), that is only allowed to mutate once per site without recombination permitted (Kimura 1969; Kimura 1971). Under this model, an estimation of neutral genetic diversity, so the expected genetic diversity (π) found when the level of new mutations balances their loss due to drift, can be described as $4N_e\mu$ the product of the effective population size Ne and the mutation rate μ (Kimura 1971). The effective population size (Ne) is defined as the number of individuals in a theoretically ideal population having the same magnitude of random genetic drift as the actual population (Hartl and Clark 2007). In an ideal population, defined as a population with balanced sex ratio, free mating, a constant number of breeding individuals per generation and an equal chance for the reproduction of each, the effective population size Ne equals the total number of individuals in the population (census population size or Nc).

In contrast, a population that does not fit these criteria Nc is usually larger than Ne (Kliman et al. 2008). However, following the assumptions of the ISM, the genetic diversity, a function of the site frequency spectrum (SFS), in an ideal population can be also calculated using the number of segregating sites S ($\theta = S/a$; $a = 1 + 1/2 + 1/3 + \ldots + 1/(n-1)$) as well as the number of nucleotide mismatches ($\pi = \theta$ = Total number of nucleotide mismatches / Total number of pairwise comparisons) (Kimura 1968; Watterson 1975). In 1989 Tajima proposed that the difference (D) between the two estimates of nucleotide diversity θ and $\theta\pi$ ($D = \pi - \theta$), could be used to test if the ISM fits the tested dataset (Tajima 1989). The rationale of his statistical test, also called Tajima's D, was that calculations of nucleotide diversity based on the number of pairwise differences (π) and the number of segregating sites (θ) differ under the impact of evolutionary processes such as selection or demographic events. That is because, in contrast to the number of pairwise differences, the number of segregating sites is indifferent to the relative frequencies of the polymorphic nucleotides at a site (Hartl and

Clark 2007). Tajima's D values for single loci that are approximately zero suggest neutrally evolving sequences whereas significant deviations from zero may suggest purifying or positive directional selection (D<0) and balancing selection (D>0) (Nei et al. 2010). When applied on a genome-wide scale, Tajima's D is also sensible to the effects of population structure and demographic changes such as population size bottlenecks or expansions (population size changes) (Maruyama and Fuerst 1985; Tajima 1989; Simonsen et al. 1995; Kreitman 2000; Whitaker et al. 2005; Hartl and Clark 2007). Thus, recent population size expansions may cause an excess of low-frequency alleles to lead to a shift of the SFS and thus to a genome-wide negative Tajima's D value (Tajima 1989; Slatkin 1996). Bottlenecks, on the other hand, lead to a reduction of the rare alleles to changes of the SFS, which lead to in genome-wide positive Tajima's D values (Tajima 1989).

The genome-wide investigations of the genetic diversity of various lineages of the pathogen *Cryptococcus neoformans* indicated that past demographic events in two of the lineages (Desjardins et al. 2017). In populations of the Fusarium head blight pathogen *F. graminearum*, genome-wide calculation of Tajima's D was performed to determine the type and strength of selection operating on regions that were in recombination hotspots (Talas and McDonald 2015). They analyzed 213 strains from several wheat fields in Germany and generated a population genomic dataset based on genome-wide SNPs based on RADseq. By performing an analysis of molecular variance (ANOVA) on the dataset, they found different results compared to a previous study and suspected an influence by the number of markers. Further, they found that the tested population of *F. graminearum* shows signs of frequent recombination by testing the population genomic dataset for linkage disequilibrium (LD) (Talas and McDonald 2015).

LD is the non-random association of alleles along a chromosome (Hartl and Clark 2007). Because it is affected by random genetic drift, the strength of observed LD depends on the size of the population (Zöllner 2008). Consequently, genome-wide measures of LD can be affected by population size bottlenecks and population substructures (Charlesworth et al. 2003; Haddrill et al. 2005). During population size bottlenecks the effects of drift may lead to elevated rates of LD inside a population (Peltonen et al. 1995; Hartl and Clark 2007). However, a recent admixture of populations with high and low rates of LD can also lead to the observation of elevated LD in the admixed population (Hartl and Clark 2007). Due to high rates of recombination and

increasing distance between markers, LD is expected to decay rapidly (Hartl and Clark 2007). This decay can be visualized using the parameter r^2 introduced by Hill and Robertson (Hill and Robertson 1968) and improved by Ohta and Kimura (Ohta and Kimura 1971) represents the strength of LD.

A high amount of LD can also be a sign of clonality, as it could be shown for the rice blast pathogen *M. oryzae* (Gladieux et al. 2018). In this study, four populations from several global areas were tested for LD decay over distance to assess their amount of recombination. It is crucial to collect knowledge about clonality in the population because clonality violates many of the common assumptions of population genomics (Grünwald et al. 2017).

Beside genome-wide effects such as population structure, demographic events such as genome size expansions or bottlenecks and random genetic drift, LD may also be locally affected by natural selection (Hartl and Clark 2007). Selective sweeps, defined as events that increase the frequency of favorable alleles in a population due to selection, have been shown to produce LD in limited regions (Prezeworski et al. 2005; Maynard Smith and Haigh 2008). Increased LD is generated by fixing not only the beneficial allele but also the stimulating region in a selective sweep (Parsch et al. 2001; Kim and Stephan 2002; Sabeti et al. 2002; Wootton et al. 2002).

Although selection sometimes causes changes in genetic diversity and LD these signatures are usually relatively weak (Hartl and Clark 2007). Natural selection subdivides into two main types called purifying (eliminates deleterious mutations) and positive selection (favors advantageous mutations). However, positive selection subdivides into balancing (maintains polymorphism) and directional selection (tends towards fixation of an advantageous allele) (Hughes 2007). The neutral theory of molecular evolution supposes purifying selection to be the primary type of selection that acts on evolving DNA sequences, while positive selection should be rare (Kimura 1968).

Inference of natural selection

Comparative population genomics offers various methods that use differences in the proportion of synonymous and nonsynonymous substitutions as a measure for selective pressures (Raffaele et al. 2010; Grandaubert et al. 2017). The difference between synonymous and nonsynonymous substitutions is if they alter the amino acid sequence the gene translates. In contrast to nonsynonymous substitutions, synonymous substitutions do not alter the amino acid sequence (Hartl and Clark 2007). The parameter ω,

a measure of selective pressures, is defined as the ratio of non-synonymous and synonymous fixed differences between species (dN/dS) (Yang and Nielsen 1998). A value of $\omega > 1$ implies positive selection, whereas $\omega > 1$ and $\omega = 1$ indicate purifying selection and neutral evolution (Yang and Bielawski 2000). Unlike previous approaches based on counting pairwise differences, modern approaches use maximum likelihood and Bayesian methods. Since in functionally essential genes only a small part of the non-synonymous are retained by natural selection, since otherwise they usually have negative fitness effects, calculations of the gene-wide average of the dN/dS offer only a small statistical power (Nielsen and Yang 1998). Therefore, methods have been developed that can calculate diversifying selection for single sites. Some of these methods allow the calculation of ω per site / codon (site models), by branch of a phylogeny (branch site) or for a few sites of predefined branches (branch-site model) (Nielsen and Yang 1998; Yang 1998; Yang and Nielsen 1998; Yang et al. 2000; Yang and Bielawski 2000). Site models based on a maximum likelihood approach allowed the comparisons of models with and without positive selection by performing a likelihood ratio test (LRT) (Goldman and Yang 1994; Muse and Gaut 1994). If a model that allows positive selection shows a higher probability the tested gene is assumed to be positively selected. It may also be advised to filter the results according to their false discovery rate (FDR) to perform a correction for multiple testing, that allows the reduction of false positives (Anisimova et al. 2007). The parameter ω can also be estimated for each branch in a phylogeny (Aguileta et al. 2009). Calculations of dN/dS can also be used on population data (Anisimova et al. 2001). Since dN/dS based methods were initially developed for use with divergent species, this is only possible with restrictions (Kimura 1977; Goldman and Yangf 1994; Muse and Gaut 1994). For example, when applied to individual populations, $\omega < 1$ can also indicate a positive selection, as the rapid fixation of alleles by selective sweeps, can produce signatures of purifying selection (Nielsen and Yang 1998). However, the use of two divergent populations has returned to reliable results (Nielsen and Yang 1998).

An analysis based on multiple genome alignments of 13 isolates of the wheat pathogen *Z. tritici* and isolates of *Z. ardabiliae*, *Z. brevis,* and *Z. pseudotritici* applied dN/dS ratio testing using the PAML package (Grandaubert et al. 2017). By using codon models 786 genes could be detected to be positively selected in *Z. tritici*, branch-wise testing using multiple genome alignments between the species revealed enrichment of positively selected effector-like proteins in *Z. tritici* and *Z. pseudotritici* but not in *Z.*

ardabiliae and *Z. brevis*. A package commonly used for the detection of ω is PAML (Yang 1997; Yang 2007). It is essential to mention that the calculation of ω may be highly biased by recombination. The presence of recombination may violate some of the central assumptions of packages like PAML which assume a constant phylogeny along the genome. Therefore, it might not be advised to use this approach for highly recombining species (Aguileta et al. 2009).

Another commonly used method for the detection of positive selection is the McDonald-Kreitman (MK) test (McDonald and Kreitman 1991). Similar to dN/dS this analytical framework relies on the difference between non-synonymous and synonymous substitutions. An essential characteristic of the MK test is that the dN/dS ratio is not only calculated for polymorphisms in species (Pn/Ps) but also those between species (Dn/Ds). If the tested locus would follow the neutral theory (Kimura 1968), Dn/Ds should be roughly equal to Pn/Ps. However, when some non-synonymous substitutions become fixed because of selection, a sign of adaptive evolution would be a ratio of Dn/Ds < Pn/Ps (Eyre-Walker 2006). Based on the equations of McDonald and Kreitman (1991) Nick Smith and Adam Eyre-Walker (Smith and Eyre-Walker 2002) could show that also the proportion of adaptive mutations (α) can be calculated using the equation $\alpha = 1 - (DsPn/DnPs)$. If a locus evolves neutral α is expected not to deviate significantly from zero, whereas a positive value α indicates adaptive evolution. Another way of quantifying deviations from neutrality of test loci is the neutrality index (NI), calculated as NI = (Pn/Ps)/(Dn/Ds) (Rand and Kann 1996). Neutrally evolving loci show a NI of approximately one, whereas NI > 1 may be an indication of purifying selection and NI < 1 for positive selection (Meiklejohn et al. 2007).

Between species genome comparisons

Comparative genomics can use genome data from different species to retrieve information about sequence conservation of shared sequences, the presence of unique sequences and gene content as well as signatures of natural selection (Hardison 2003). Using comparative genomics between genomes of the smut fungi *U. maydis* and *Sporisorium relianum* regions with low sequence conservation could be detected (Schirawski et al. 2010). The authors reported 43 regions with higher sequence diversity and enrichment in effectors. Jackson and co-workers compared whole genome sequences of the fungal plant pathogens *C. albicans* and *C. dubliniensis* (Jackson et al. 2009). Since these two pathogens have very different rates of virulence, the authors were interested in species-specific genes. In total 168 species-specific genes were

detected. Because of the different gene repertoires, the authors suggested that the less virulent pathogen *C. dubliniensis* has lost specific essential pathogenicity genes (Jackson et al. 2009). Using comparative genomics also more significant species-specific regions can be detected. By comparing the genomes of *F. graminearium*, *F. verticilloides* and *F. oxysporum f. sp. lycopersici*, lineage-specific regions (LS) were detected (Ma et al. 2010). These regions were also proven to induce virulence in a former avirulent strain of *F. oxysporum*. This Ph.D. thesis uses methods of comparative genomics to investigate the influence of species-specific genes in the *Zymoseptoria* species complex.

1.4. Generating population genomic datasets by collecting genetic variation

In general, studies that apply population genomic methods on datasets containing genome-wide SNPs commonly use a short-read-alignment approach against a reference genome (i.e., Fawcett et al. 2014; Schmidt et al. 2015; Talas and McDonald 2015; Cissé et al. 2018).

In contrast to this most commonly used procedure (Figure 3), this study employs an approach using multiple genome alignments (MGAs) generated with *de novo* whole genome assemblies of several *Zymoseptoria* species (Figure 3). Aligning short sequencing reads against a single reference genome is generally limited by repetitive elements, that are harder to align and therefore produce a higher number of mismatches and unaligned reads (Figure 4 A) (Treangen and Salzberg 2012). Not only repetitive sequences have an impact on reading alignment. Structural variations and large rearrangements such as indels (insertions and deletions), inversions, translocations or duplications also produce distinct patterns of discordantly mapped reads (Abel and Duncavage 2015). Structural variations and their influence on virulence have already been addressed in many model systems of fungal plant pathogens (De Jonge et al. 2013; Jones et al. 2014; Plissonneau et al. 2016). Building on this fact, several methods have been developed which use read-depth, read-pairs, split-reads or assemblies to find structural variations (Alkan et al. 2011). Many software for the detection of structural variation like CNVer (Medvedev et al. 2010), DELLY (Rausch et al. 2012), LUMPY (Layer et al. 2014) or HYDRA (Quinlan et al. 2010) combines multiple of these methods.

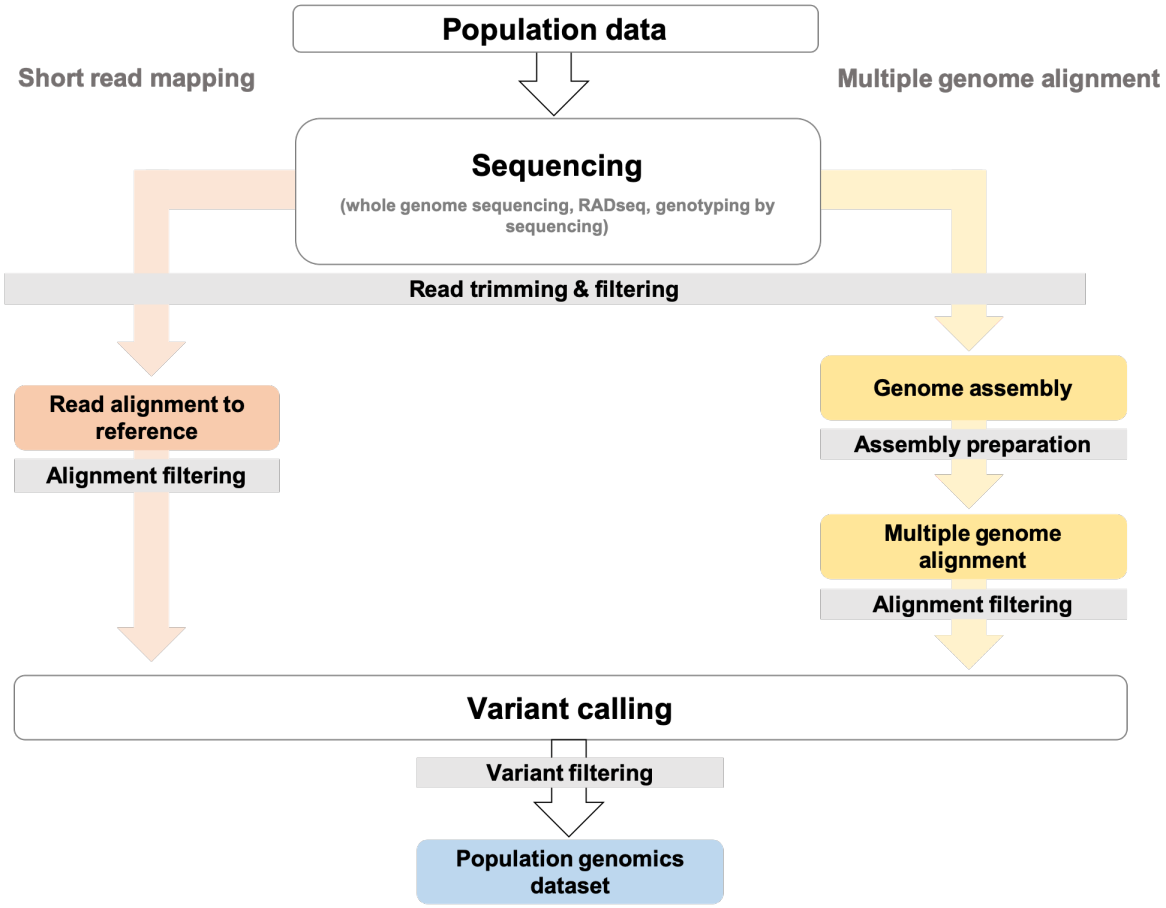


Figure 3 - Illustration of two different pipelines for generating SNP datasets. Short read alignments (red), as well as multiple genome alignments, can be used as a starting material to create a dataset for population genomic analysis. The preparation of such a data set requires in both methods wholly sequenced genomes of several individuals of a population. Depending on the genome size, whole genome sequencing (WGS) or representation methods such as RADseq or genotyping by sequencing (GBS) can be used. The sequencing reads must be filtered after successful sequencing to reduce the bias to the later analysis. However, from this point, further procedures of the methods differ. The red path shows that nucleotide variants can be called directly from short read alignments. The method, which is based on multiple genome alignments (yellow path), requires an intermediate step before the alignment in which the short sequencing reads genome are assembled. Finally, from the multiple genome alignments of the assemblies, the nucleotide variants are called.

The rapidly increasing number of whole genome sequences of entire populations led to a concept called pangenome (Tettelin et al. 2005). The pan-genome is defined as the entirety of the core genome (present in all individuals) and the dispensable genome of a species (Medini et al. 2005; Vernikos et al. 2015; Marschall et al. 2018). Pan-genome analyzes have been used extensively in the past to investigate the influence of species-specific genes on the speciation of fungal plant pathogens (Liu et al. 2014; Upadhyaya et al. 2015; Walkowiak et al. 2016). A recent study on the pangenome of the plant pathogenic ascomycete *Z. tritici* reported 6600 accessory genes (Plissonneau et al. 2018). However, *Z. tritici* not only shows accessory gene content the genome of

this pathogen is also structured in core and accessory chromosomes and shows a high degree of plasticity (Wittenberg et al. 2009; Goodwin et al. 2011).

Figure 4 shows how different presence-absence polymorphisms affect population genomic analysis based on short read and multiple alignments. However, working with multiple genome alignments has some advantages over using short read alignments. Both methods make it possible to detect connections (links) between contigs or scaffolds that are not in the reference. Also, both approaches allow finding inversions of sequence sections (Figure 4 A, B1). However, multiple genome alignments in their non-projected (Figure 4 B1) form contain further useful information about regions that are not present in the reference genome due to insertions or deletions. These can easily be examined by projecting against another genome of the genome without creating a new dataset (Figure 4 B2). However, this option was not used in this thesis because of the low sample size.

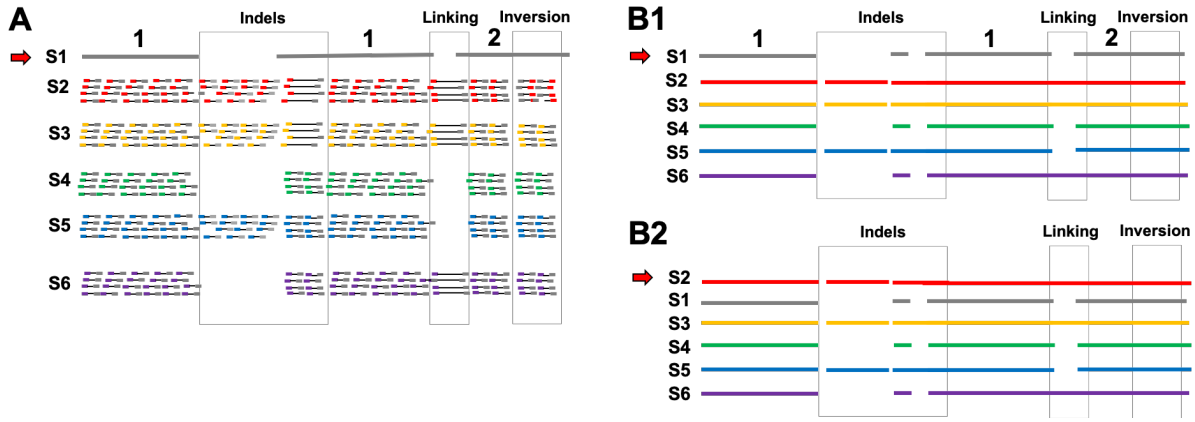


Figure 4 – Example schemes illustrating the differences in structural variant detection between short read and multiple genome alignments. The figure shows how different presence-absence polymorphisms affect population genomic analysis based on short read and multiple alignments. Thus, both methods can demonstrate links between contigs or scaffolds separated into the reference genome, or inversions. One of the strengths of multiple genome alignments, however, is the possibility to study sequences that are not present in the reference genome due to presence-missing polymorphisms by projecting them against another genome. A) Scheme of a short read alignment against two reference chromosomes. B1) Scheme of multiple genome alignments against the same two reference chromosomes. B2) Multiple genome alignments (MGAs) projected against another reference to illustrate the benefit of MGAs for analyzing sequences with presence-absence polymorphisms. Indel = Insertion or Deletion; Linking = Reads provide information about a link between chromosome 1 and 2 in some sequences.

Genome sequencing

Advances in the effectiveness in combination with price reductions of short-read sequencing technologies like Illumina, 454 pyrosequencing or Ion Torrent made the sequencing of entire genomes affordable (Shokralla et al. 2012; Bokulich et al. 2013). However, sequencing of large may still be too expensive so that the use of reduced-representation sequencing Examples for such as restriction site-associated DNA sequencing (RADseq) or genotyping by sequencing (GBS) might become necessary (Baird et al. 2008; Elshire et al. 2011). Many fungal plant pathogens contain relatively small genomes (18 - 180 Mb; Table 1) which allow for sequencing directly in full length with high coverage. Additionally, most filamentous fungi are predominantly haploid (Nieuwenhuis and James 2016).

There are several unknown dimensions of sequencing errors that can result in inaccurate genome assemblies, alignments or alignments leading to bad quality data. High sequencing coverage can ameliorate these problems by facilitating the identification of sequencing errors from natural actual biological variation. *De novo* genome assemblies that rely on high coverage sequencing reads is an example of a method that reproducibly distinguishes sequencing error from real biological information (Edgar and Flyvbjerg 2015). Next, to sequencing errors, any remaining adapter sequences can also profoundly impact assembly quality by introducing mismatches during the short read alignment procedure (Edgar and Flyvbjerg 2015). Consequently, filtering raw sequencing reads from positions with bad quality and remaining adaptor sequences could increase the assembly quality noticeably (Bolger et al. 2014).

In contrast to high throughput technologies like Illumina or 454 pyrosequencing, the single molecule real-time sequencing technology (SMRT) developed by Pacific Biosciences has a considerably lower throughput but produces considerably longer and highly accurate DNA sequences from individual unamplified molecules (Eid et al. 2009). Longer read lengths of several thousand base-pairs and improvements in sequencing error reduction (when using coverage depth of > 8x) that came with the new PacBio RS II platform make this sequencing technology an excellent tool to generate data for *de novo* genome assemblies of small genomes (Roberts et al. 2013). In this study SMRT sequencing technology is used for the generation of high quality *de novo* genome assemblies that serve as reference sequences for the sister species of *Z. tritici*.

Calling variation from short-read alignments

Short read alignment is usually performed using software like bowtie (Langmead et al. 2009), bowtie2 (Langmead et al. 2009) or bwa (Li and Durbin 2009; Li and Durbin 2010). For the first evaluation of the alignment result programs like Qualimap (García-Alcalde et al. 2012; Okonechnikov et al. 2015) can be useful. Subsequent filtering of the mapped read data is crucial for reducing the amount of falsely aligned reads. Much modern variant-calling software like Samtools (Li et al. 2009; Li 2011) already account for errors specific to the sequencing technology used (Liu et al. 2013). However, the algorithms used for the calculation of sequencing errors are highly sensitive to read depth, which means that it is advised focusing on loci, that show a minimum read depth of 20x (Andrews & Luikart 2014). Another general problem with short read alignment procedures is the occurrence of PCR duplicates. Here approaches based on reduced genome representations like RADseq seem to be more affected by this problem because of the usage of unified PCR primers (Davey et al. 2013; Hohenlohe et al. 2013). It is also recommended to remove reads that belong to paralogous regions in the genome. To do this a reference genome is needed to mask paralogs and read contaminations (Haussler et al. 2009; Baxter et al. 2010; Grünwald et al. 2017). Several tools like Samtools (Li et al. 2009; Li 2011), Bamtools (Barnett et al. 2011) or Picard (http://broadinstitute.github.io/picard/, Effective: 10.01.2018) are used to manipulate and filter short read alignments. These programs allow users to remove reads with low alignment quality, duplicates, paralogs or unmapped reads from the dataset.

Variant calling can be performed using software like GATK (McKenna et al. 2010) or Samtools (Li et al. 2009; Li 2011) and many others (i. ex. compared in Liu et al. 2013). However, programs like Stacks (Catchen et al. 2011; Catchen et al. 2013) or Tassel (Bradbury et al. 2007) can call SNPs without any need of a reference genome. It is crucial to filter the nucleotide variants to remove false positive SNPs in the dataset. It is crucial to filter the dataset for missing data, copy number variation and minimum allele frequency (Li et al. 2009b; Andrews & Luikart 2014; Andrews et al. 2016). This filter conditions can be applied using programs like Vcftools (Danecek et al. 2011). Unfortunately, there is no comprehensive set of filter conditions available whereby it is necessary to adjust the filters to each dataset separately.

Calling variation from multiple genome alignments

The technique using multiple genome alignments consists of four sequential steps (Figure 3) comprising sequencing and read filtering, *de novo* genome assembly, alignment with subsequent filtering and variant calling for producing a population genomic dataset including genome-wide SNPs (Figure 3 right). The method used in this Ph.D. thesis and described here has already been used for population genomic analyses of the wheat pathogen *Z. tritici* (Grandaubert et al. 2017). Using multiple genome alignments for comparative and population genomics keeps more information about chromosomal rearrangements and presence-absence polymorphisms of chromosomes in and between populations. This additional information may be necessary because chromosomal rearrangements have been shown to have the potential evolutionary advantages in fungal plant pathogens. For example, rearrangements have been shown to contribute to host specialization in *Z. tritici* and *V. dahliae* (Seidl and Thomma 2014; Hartmann et al. 2017). However, whole chromosomal presence-absence polymorphisms as found for accessory chromosomes can also impact fitness of fungal plant pathogens as shown in *Z. tritici* (Habig et al. 2017), *F. oxysporum* (Ma et al. 2010) or *Nectria haematococca* (Coleman et al. 2009) and other plant pathogenic fungi (Mehrabi 2017, Soyer 2018). Since the first sequencing assembly method came up about 40 years ago the methodology steadily improved (Flicek and Birney 2010; Phillippy 2017). In 1979 a first computational method was realized in FORTRAN that allowed the assembly of DNA sequences out of sequencing reads (Gingeras et al. 1979). Today two different approaches for generating whole genome assembly are commonly used. They differ by the need of a genome reference that guides the assembly (comparative) or if they completely assemble the sequence without any knowledge about a reference sequence (*de novo*) (Pop 2009). In the absence of a reference genome, *de novo* approaches are often necessary. For those analyses, this assembly method may be superior, since it also includes sequences, which are not present in any reference. Thus, *de novo* assemblies provide additional information for population genomic approaches. Depending on the length of the sequencing reads two different techniques can be used to generate genome assemblies. Short-reads as produced by Illumina or 454 pyrosequencing runs are mainly assembled by programs using de Bruijn graphs (Pevzner et al. 2001) like SPAdes (Bankevich et al. 2012), SOAPdenovo2 (Luo et al. 2012) or IDBA-UD (Peng et al. 2012).

In contrast, long reads as generated by Sanger or SMRT sequencing require overlap based methods (Myers 1995; Eid et al. 2009) as used in assemblers like Canu (Koren et al. 2016), Falcon (Chin et al. 2016) or SMRTAssembly (©Pacific Biosciences). The application of sequencing reads with average lengths of 10kb like they are common in the example for SMRT sequencing can often span over repetitive regions such as TEs, since their length is larger than the most common repeats (Phillippy 2017). Although using recent high-throughput technologies it might be affordable to sequence entire small genomes in high coverage depths of short reads, it might still be too expensive creating decent coverage depths using long-read approaches like SMRT sequencing (Chakraborty et al. 2015; Forouzan et al. 2017). For short-read assemblies at least a coverage depth higher than 100x is optimal (Ekblom and Wolf 2014; Chakraborty et al. 2015; Forouzan et al. 2017). Whenever possible, it might be beneficial to use long-read sequencing for generating genomes *de novo*. Depending on the coverage depth obtained with long reads distinct strategies are available. Low coverage depths of long reads lead to fragmented genome assemblies, which may require the use of a hybrid method including long and short read data (Chakraborty et al. 2015). Specialized assemblers like hybridSPAdes are useful tools in such cases (Antipov et al. 2016). As a simple rule of thumb coverage depths of 25x to 35x of long-reads may require a hybrid assembly approach, whereas higher coverage depths of more than 60x enable *de novo* assembly using only long reads (Chakraborty et al. 2015).

There are several tools available like Tba (Blanchette 2004), Mugsy (Angiuoli and Salzberg 2011) or progessiveMauve (Darling et al. 2010) for generating multiple genome alignments. Tba and progressiveMauve use information about the phylogeny of the samples; therefore they require phylogenetic information a priori (Blanchette 2004; Darling et al. 2010). Since raw multiple genome alignments contain a large number of blocks with the different quality, it is recommended to re-align every aligned block again with a multiple sequence aligner like Mafft (Katoh et al. 2009) or T-Coffee (Notredame et al. 2000). This step should decrease the number of false aligned regions. After re-aligning the multiple genome alignments should be filtered to remove regions with low quality. So far there is no comprehensive set of filters available, but there are some general considerations. As a start one may be advised to remove alignment blocks that undergo a certain block length threshold of for example 1kb. It is highly recommended to adjust this threshold according to the dataset. A second applicable filter would be a threshold for block size to ensure that only alignment blocks that contain

information for all strains in the dataset remain. Because poorly aligned blocks may show a more significant number of gaps than correct blocks, one could also use a gap ratio filter, that performs window-wise calculations of gap ratios and discards windows that exceed a given threshold. Again, this threshold needs adjustment for each dataset separately manual comparison filter success. Tools like Maffilter (Dutheil et al. 2014) provide these type of filters and some more options to process multiple genome alignments. Maffilter, for example, provides the possibility of calling variable sites directly from MGAs or to calculate sequence diversity statistics such as estimators of the nucleotide diversity in sliding windows (Dutheil et al. 2014). So far multiple genome alignments have been used as a basis of population genomics in several studies. For example, performing comparative population genomics on isolates of *Z. tritici* and *Z. ardabiliae* high recombination rates and signatures of adaptive evolution could be detected (Grandaubert et al. 2017; Stukenbrock and Dutheil 2018).

2. Objectives and research questions

The extraordinary genome plasticity of the *Zymoseptoria* species complex

The wheat pathogens ascomycete *Z. tritci* shows extraordinary genome plasticity with length and presence-absence polymorphisms of whole chromosomes. Studies have shown that accessory chromosomes also occur in the wild sister species. However, it is not known to which extent genome plasticity also occurs in the wild grass associated pathogens of the genus *Zymoseptoria*. This section assesses the changes in genome structure that can be correlated with species divergence:

1) Can genome data confirm that accessory chromosomes are an ancestral trait of the genus *Zymoseptoria*?
2) Are there any differences in synteny between the *Zymoseptoria* species?
3) Can synteny breaks be associated with the position of repetitive elements?

Signatures of intra-specific genetic variation point to differences in population structure and demography

The *Zymoseptoria* species complex includes both important cereal pathogens and wild grass associated pathogens. So far, little is known about how the adaptation to different ecosystems has affected the development of population structures or demographic trends in this pathosystem. This section of the thesis addresses variation in population structure and demographic history among *Zymoseptoria* species from managed and natural ecosystems. Specifically, it will address:

1) To which extent are there differences in the genetic diversity between pathogens from managed and natural ecosystems?
2) Are there any differences in population structure observable between the *Zymoseptoria* species?
3) Do the *Zymoseptoria* species show differences in their demographic history?

Ongoing adaptive evolution and the detection of effector candidates in the genus *Zymoseptoria.*

Plant pathogenic fungi co-evolve with their host. Genes involved in adaptation to the host or the local ecological conditions are affected by natural selection. Especially co-evolving genes are often subject to positive selection, a type of selection that favor adaptive non-synonymous changes in the gene sequence. Particularly noteworthy in

this context are genes encoding effectors, a class of proteins that can manipulate the immune system and are thus part of the direct interaction between host and pathogen. This section identifies genes with potential importance in virulence:

1) Which genes show signatures of recent selection possibly reflecting adaptation to distinct hosts and ecosystem?
2) Do the pathogens of the *Zymoseptoria* species complex differ in the proportion of genes encoding effector candidates in their genome?

3. Methods

3.1. Fungal Isolates

This thesis used a group of 13 *Z. tritici,* 9 *Z. brevis* and 17 *Z. ardabiliae* field isolates from different geographical locations. All isolates of *Z. tritici* were isolated from bread wheat (*Triticum aestivum*) from various geographically distant locations collected and different time points in Germany, Denmark, Iran, France, and the Netherlands (Table 2). The isolates of *Z. ardabiliae* were collected from wild grasses of the genera *Lolium*, *Agropyron* and *Dactylis* at several locations in the province Ardabil, Iran in 2004 and 2011. *Z. brevis* was isolated from leaves of *Phalaris minor* and *P. paradoxa* in Ilam province in a region next to the city of Dehloran in Iran (2009). A more detailed overview of the isolates used in this study can be found in the supplementary material (Table S 1).

Table 2 - Overview about origin, host and collection time of the *Zymoseptoria* isolates used.

Species	Isolates #	Year	Location	Host	Published
Z. ardabiliae	4	2004	Iran	*Agropyron sp.* *Lolium sp.*	Stukenbrock et al. 2012
	13	2011	Iran	*Agropyron tauri* *Lolium sp.* *Dactylis glomerata*	Stukenbrock and Dutheil 2018
Z. brevis	9	2009	Iran	*Phalaris minor* *Phalaris paradoxa*	Quaedvlieg et al. 2011
Z. passerinii	1	1995	USA	*Horedum vulgare*	Stukenbrock et al. 2007
Z. pseudotritici	4	2004	Iran	*Agropyron tauri* *Dactylis glomerata*	Stukenbrock et al. 2012
	23	2004	Iran	*Agropyron sp.* *Dactylis glomerata* *Dactylis sp.*	Unpublished
Z. tritici	1	1995	Netherlands	*Triticum aestivum*	Kema & Silfhout 1997
	2	2004	Iran	*Triticum aestivum*	Grandaubert et al. in press
	4	2009	Denmark	*Triticum aestivum*	Grandaubert et al. in press
	1	2013	France	*Triticum aestivum*	Grandaubert et al. in press
	5	2013	Germany	*Triticum aestivum*	Grandaubert et al. in press

3.2. Molecular methods

3.2.1. Preparation of fungal cell cultures

Fungal strains were inoculated on solid YMS medium directly from glycerol stock (-80°C) and incubated for five days at 18°C. Subsequently, fungal cells were transferred to culture flasks containing 25mL YMS liquid medium using a pipet tip. The liquid cultures were incubated in a shaking incubator at 18°C with 200rpm for five days. Media and Buffers used in this thesis can be found in the supplement (8.1).

3.2.2. Karyotyping using pulsed-field gel electrophoresis

Previous studies indicate the presence of accessory chromosomes in *Z. ardabiliae*, *Z. brevis*, *Z. tritici* and *Z. pseudotritici* (Stukenbrock et al. 2007; Stukenbrock et al. 2011; Eschenbrenner 2013). However, not much is known about the genome composition of the wild grass associated with *Zymoseptoria* species. This part of the thesis aims to characterize complete genome karyotypes of various isolates of *Z. ardabiliae*, *Z. brevis*, *Z. tritici,* and *Z. pseudotritici.*

Chromosome plugs for pulsed-field gel electrophoresis (PFGE) were generated using cell cultures grown as described previously. Mature cell cultures were filtered using Miracloth (Merck, Darmstadt, Deutschland) filter tissue to remove any hyphal cells and retrieve a suspension of single cells. Using cell suspensions with single cells maximizes lysis success. The single cell suspensions were pelletized with 3500rpm for 10min and resolved in 20mL Tris-HCl (pH 8). This step was repeated once, and cell dilutions (1:100) were counted using a Neubauer counting chamber (Brand, Wertheim, Germany). In total $5x10^8$ cells/mL were embedded in 1.1% low range agarose (Bio-Rad, Munich, Germany) in 0.5x TBE buffer, poured in into plug casting molds and solidified for 1h at 4°C. Solidified agarose plugs were incubated in 5mL lysis buffer at 55°C for 48h with an exchanging of lysis buffer after 24h. Subsequently, the gel plugs were rinsed 2-3 times with 1x TE buffer for 20min until all remaining proteases, SDS and cell debris were removed. For long time storage gel plugs were kept in 0.5M EDTA pH 8 at 4°C.

In the PFGE method, electric fields were applied at a voltage of 5V / cm at an angle of 120° diagonally to the position of the agarose gel. The relative position of the active electric field to the agarose gel was changed in defined time intervals (Figure 5). The

different position of the pulsating electric fields led to a regular change of direction of the chromosomes on their way through the agarose gel (0.8% agarose). Smaller chromosomes passed faster through the mesh of the agarose gel, allowing for size separation (Herschleb et al. 2007). The duration of the pulse was linearly increased during the experiment from 50 seconds to the start time up to 150s in the end. As a running buffer, 0.5x TBE was pumped through the system at a constant temperature of 13 °C with 80% pumping force. The size standard used was a ladder based on the genome of *Saccharomyces cerevisiae*.

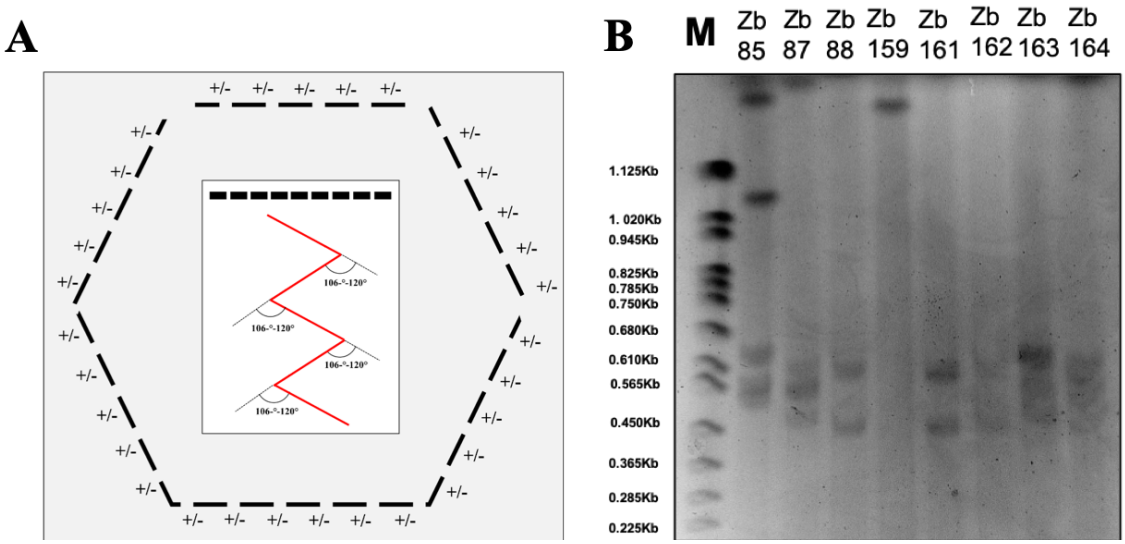


Figure 5 - Schematic illustration of pulsed-field gel electrophoresis (PFGE). A) Red lines indicate the route a DNA molecule takes in the pulsing field with a fixed field angle between 106° to 120°. The pulsing field leads to periodically re-ordering of DNA molecules in the agarose gel, leading to directional changes that make it easier for large molecules to pass the agarose mesh. The fixed field-angle that adjusts the orientation of the electric field equals the angle of the directional change in molecule passage. Picture modified after Herschleb et al. 2007 and the manual of the CHEF-DR III Chef Mapper (Bio-Rad, Munich, Germany). B) Example of a PFGE gel. The picture shows the separation of chromosomes of less than 2.2Mb in *Z. brevis* using the PFGE on a CHEF-DR III Chef Mapper (Bio-Rad, Munich, Germany). M = Marker (*S. cerevisiae*; BioRad, Munich, Germany).

After electrophoresis, all gels were stained for 30min in ethidium bromide staining solution (1 µg/mL ethidium bromide in H_2O). After washing with H_2O once, band patterns were detected with the GelDocTM XR+ system (Bio-Rad, Munich, Germany).

3.3. Bioinformatical methods

All scripts that were programmed in the course of this thesis are included on the included USB stick (8.15). Italic letters mark the names of the scripts as well as the parameters and functions of the software.

3.3.1. Whole genome sequencing data

When this project started several genomes of *Z. tritici* (8), *Z. ardabiliae* (13) and *Z. brevis* (8) were already sequenced using a HiSeq2500 Illumina platform (AROS Applied Biotechnology, Denmark). Short read sequencing data and genome assemblies of two *Z. tritici* isolates (IPO323 as Zt09 and Zt11 as STIR01A48b), and one isolate of *Z. ardabiliae* (Za17) used in this thesis were published previously (Goodwin et al. 2011; Stukenbrock et al. 2011). However, in this thesis, single-molecule real-time (SMRT, © PacBio) sequencing was used to generate improved assemblies for reference isolates of *Z. ardabiliae* (Za17), *Z. brevis* (Zb87), *Z. passerinii* (Zpa63) and *Z. pseudotritici* (Zp13), as well as for two additional *Z. tritici* isolates (Zt05 and Zt10). The latter two are published in Haueisen et al. 2017. The SMRT sequencing was done at the Max Planck Genome Center in Cologne, Germany (*Z. tritici*, *Z. passerinii*) and the ETH Genome Center, Zürich, Switzerland (*Z. ardabiliae*, *Z. brevis*, *Z. pseudotritici*).

3.3.2. Building assemblies *de novo*

Assembly of short read sequencing data

Previously to this work, *de novo* genome assemblies of 12 *Z. tritici*, 8 *Z. brevis*, and 27 *Z. pseudotritici* isolates were prepared by co-workers using SOAPdenovo2 (Luo et al. 2012; Grandaubert et al. 2017). In a master thesis, which was supervised in the context of this thesis, the genomes of *Z. ardabiliae* were re-assembled *de novo* using the software SPAdes v3.7 with standard settings (Bankevich et al. 2012; Braker 2017). Summary statistics about the assemblies were generated using Quast v4.6.3 (Gurevich et al. 2013). Finally, all scaffolds with a length less than a threshold of 1 Kb were removed from the *de novo* assemblies using a custom python script ("*AssemblyFilter.py*").

For retrieving the average coverage depth after *de novo* assembly, the raw sequencing reads were filtered and trimmed using Trimmomatic (Bolger et al. 2014). In this context, positions of bad quality at both ends of the read were removed by cutting of bases with a quality below the threshold ("*LEADING:3*" and "*TRAILING:3*"). Additionally, using a

sliding windows approach ("*SLIDINGWINDOW*"), the read was cut once when the average quality of a window (4bp in size) fell below a given threshold (20). Further, using the parameters "*ILLUMINACLIP*" with TrueSeq2-SE adaptors as a template and "*MINLEN*" potentially remaining adaptors and reads with a minimum length less than 50bp were removed from the dataset. Next, using standard settings, filtered reads were mapped against their respective reference genome assembly and saved the result in SAM format using Bowtie2 v2.2.9 (Langmead et al. 2009). The software Samtools 1.3.1 (Li et al. 2009) was used to convert SAM format into binary format (BAM) ("samtools *view*") using Samtools. Subsequently, the same software sorted was used with standard conditions to sort ("samtools *sort*") and index ("samtools *index*") the BAM files. Finally, the mean coverage of each *de novo* assembly was calculated using BamQC a software implemented in Qualimap v2.2.1 (Okonechnikov et al. 2015) with standard conditions. The pipeline for the coverage calculation was included in "*CoverageCalc.py*".

Assembly of long read sequencing data

Genome assemblies of Zb87, Zt05, and Zt10, were generated *de novo* and error corrected using the HGAP3 (Chin et al. 2013) protocol and Quiver, both included the software SMRTAnalysis 2.3.0 (© PacBio) However, the genomes of Za17 and Zp13 were prepared by a collaborator using the same software (Daniel Croll unpublished). Additionally, a *de novo* genome assembly of the *Z. passerinii* isolate Zpa63 was generated using the HGAP4 protocol which is included in the recent version of the SMRTAnalysis software, a part of SMRT® Link v5.1.0 (© PacBio). Both programs were used with default settings in exception of a manual adjustment of the seed read length used during the assembly procedure. Each assembly process started with default settings and a seed read cutoff set to "Auto" mode. The minimum length determined the seed read length for reads that were used as references for the pre-assembly step (Figure 6).

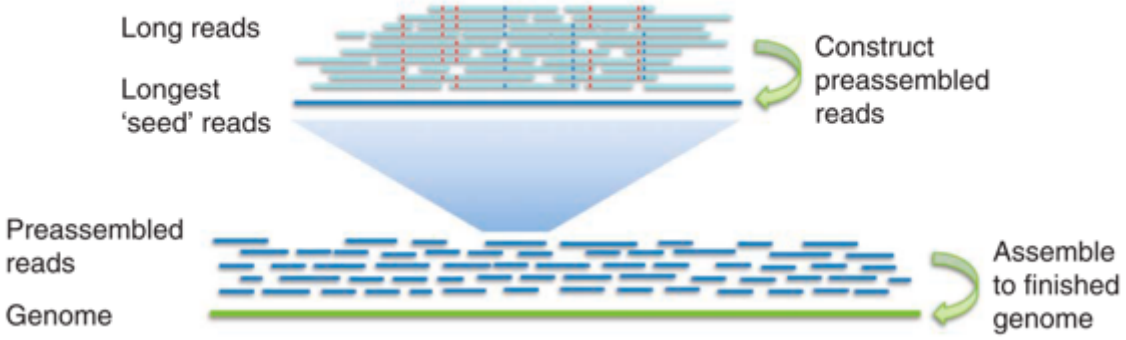


Figure 6 – Hierarchical genome assembly process (HGAP) protocol for the generation of the *de novo* genome assemblies from long sequencing reads. The longest reads serve as the reference for a preassembly and are called seed reads. The final step of the pipeline assembles the preassembled reads into the final genome assembly (Chin et al. 2013).

The result was refined by choosing estimated seed values as a starting point for further assembly processes. Thus, the length of the seed was reduced or increased by 1000-2000bp, and the assembly process was repeated. For the quality assessment the assembly size, the number of contigs, the N50 value, and the number of ambiguous nucleotides (N) per 100kb were calculated using the software Quast v4.6.3 (Gurevich et al. 2013). In cases when increasing or decreasing the seed length resulted in better assembly quality, additional assemblies were made using seed read lengths that were reduced or increased. This procedure was continued until the quality of the resulting assemblies decreased again. Among the assemblies that were generated by this procedure, the assembly statistics were compared to choose the best version. The ideal version of the assembly should have the highest N50, the smallest number of components, and the largest size of all versions produced. If none of the assemblies met all of these criteria, the following ranking of features was applied: N50> the number of contigs> size. An illustrated explanation of this procedure can be found in the supplement (Figure S 1).

3.3.2.1. Detection of repetitive genome content

Repetitive elements

All synteny analyses were performed using repeat masked *de novo* assemblies of Za17, Zb87, Zp13, Zpa63, Zt05, Zt09, and Zt10. Repetitive sequences are known to affect genome alignments by generating miss-alignments (Frith et al. 2010). Thus, to reduce this bias on the synteny analyses, repeats were removed from the *de novo* assembly by masking. No annotation of repetitive elements was available for the tested genomes. Consequently, the repetitive genome content was annotated *de novo* using the programs RepeatModeler version open-4.0.7 (Smit and Hubley 2018) and RepeatMasker version open-4.0.7 (Smit et al. 2013). In phase one, I of the pipeline (Figure 7) the *de novo* genome assemblies were screened for repetitive elements using the software RepeatModeler. Using BuildDatabases with 30 processes ("*-pa 30*") the genomes were converted into a database compatible with the NCBI search engine for repetitive element screening using BuildDatabase. Subsequently, the repetitive contents of the genomes were annotated using RepeatModeler *("-engine ncbi*").

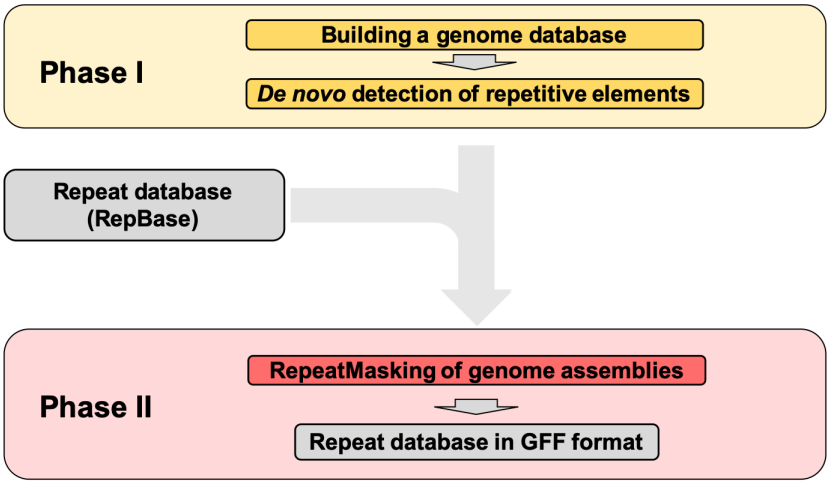


Figure 7 - Pipeline for the detection of repetitive genome content. The pipeline consists of two parts. In the first part, repetitive elements were detected *de novo*. The data set thus generated is associated with the information of a public database. In part two, the repeating part of the genome is masked with the combined data set. Grey boxes indicate input data and output data, whereas differently colored boxes indicate analyses and other processing.

For the preparation of the next phase, a dataset containing repetitive elements was generated by merging the previous results of the *de novo* detection of repetitive sequences with fungal repeats downloaded from RepBase 23.05 (Bao et al. 2015). In Phase two all entries in the previously modified fungal repeat database ("*-lib*") were

annotated in the genomes using the software RepeatMasker. In the final step, the data was converted into the gene feature format (GFF) to ensure compatibility with the software used later in this study.

Telomeric repeats

As an indicator for assembly quality the *de novo* genome assemblies of Za17, Zb87, Zp13, Zpa63, Zt05, and Zt10 were screened for the distribution of telomeric repeats (Figure 8).

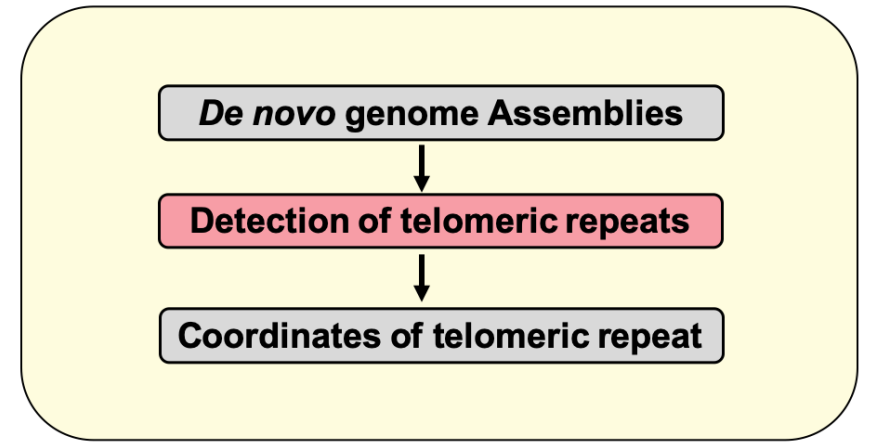


Figure 8 - Pipeline for the detection of telomeric repeats in genomes. This pipeline detected the genome-wide distribution of a common telomeric repeat to investigate the presence of fully assembled chromosomes. Grey boxes indicate input data and output data, whereas differently colored boxes indicate analyses and other processing.

The *de novo* assemblies were screened for the telomeric repeat motif CCCTAA ("-*pattern '<N(0,6)CCCTAACCCTAA*' ") and the complementary pendant TTAGGG ("-*complement*") using the program Fuzznuc included in the Emboss package (Rice et al. 2000). The sequence pattern used here is a common human-type telomeric repeat common among many vertebrates (Meyne et al. 1989; Fulnečková et al. 2013) and had been successfully used for the detection of telomeres in several plant pathogenic fungi like *F. graminearum* (King et al. 2015) or *F. circinatum* (Wyk et al. 2018).

Visualization of repetitive genome content

The repetitive genome content of Za17, Zb87, Zp13, Zpa63, Zt05, and Zt10 was visualized as circular ideograms (Figure 9) using the software Circos (Krzywinski et al. 2009) as part of a custom python script ("*RunCircos_GenomeRepresentation.py*").
An essential requirement for generating an ideogram using Circos was to provide a plain-text file that defines the karyogram ("*karyotype*"). Consequently, chromosome

names and sizes were extracted directly from whole genome assemblies and converted them into a plain-text format compatible with the software using a custom python script ("*KaryogramFromFasta.py*"). Next, density information of the telomeric and repetitive data was added to the circle plots as histograms ("*type* = *histogram*"). The density information for the tracks was generated using sliding window approaches ("*Repeats2Circos.py*" and "*Telomeres2Circos.py*"). The number of complete chromosomes was calculated based on the distribution of the density of the telomeric repeats ("*CompleteChromosomes.py*"). For this, a window-wise approach was used to calculate the telomere hits used for the representation with Circos and counted the number of windows with more than ten hits of telomeric repeats. A contig was considered as complete as long it harbored one or more windows fitting that threshold.

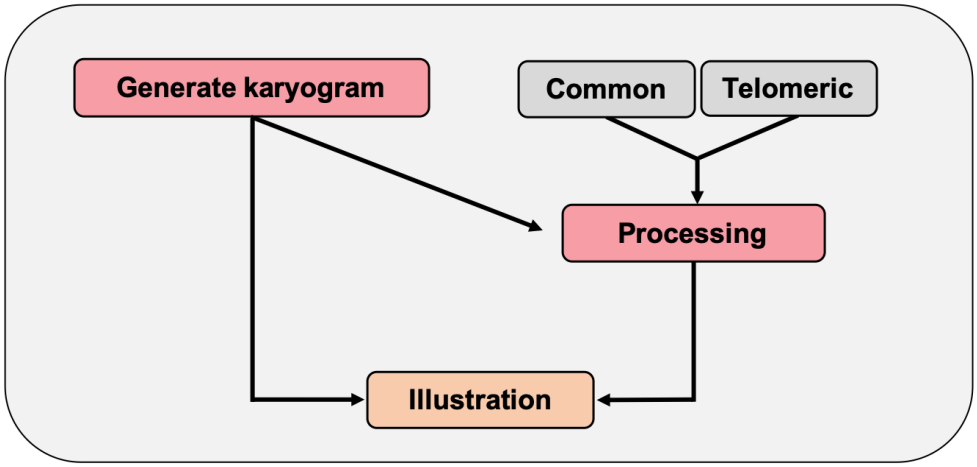


Figure 9 - Pipeline for the circular illustration of repetitive genome content. This pipeline produced graphic illustrations of the *de novo* genome assemblies generated based on SMRT sequencing. For this purpose, ideograms of the genome were created, which contain both density information of the repetitive element content and the telomeric repeats. Grey boxes indicate input data and output data, whereas differently colored boxes indicate analyses and other processing.

3.3.3. Analyzing synteny using pairwise genome alignments

In this study, pairwise comparisons of *de novo* genome assemblies were used to analyze the conservation of synteny between the members of the *Zymoseptoria* species complex (Figure 10). In the first step of the pipeline, the previously collected repeat information (3.3.2.1) was used to mask all repetitive genome content in the reference genomes of Za17, Zb87, Zpa63, Zp13, Zt05, Zt09 and Zt10 (3.3.3.1). Subsequently, the initial data set was generated from the pairwise genome alignments between the repeat-masked reference genomes. This dataset provided insight into the synteny

between the *Zymoseptoria* by the visualization of the pairwise genome alignments (3.3.3.2). Further, the data set enabled the detection of synteny blocks (3.3.3.3) and in consequence the definition synteny breakpoints with subsequent correlation studies between breakpoints and the distribution of repetitive elements (3.3.3.4).

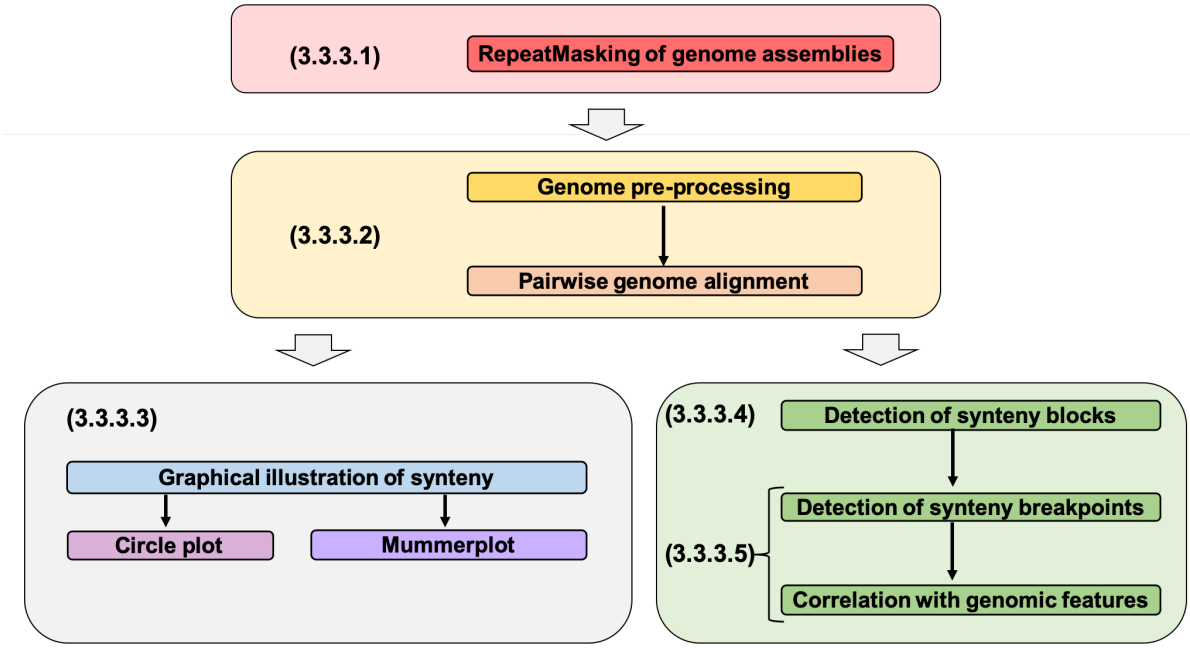


Figure 10 - Pipeline for the analysis of genome synteny in the genus *Zymoseptoria*. This graph illustrates the complete pipeline created to study the synteny in the *Zymoseptoria* species complex. It consisted of repeat masking (red), pairwise genome alignment (yellow), the evaluation of graphical illustrations of the synteny (grey) as well as the detection of synteny blocks and breakpoints (green). The numbers in brackets refer to the corresponding sections in the method part of this thesis.

3.3.3.1. Genome pre-processing and pairwise alignment

First, the coordinates for each possible combination of pairwise genome alignments were generated and collected (Figure 11).

The repeat databases generated in 3.3.2.1 were converted into browser extensible data (BED) format ("*Repeats2bed.py*"). Subsequently, Maskfasta a tool implemented in the software package Bedtools v.2.27.1 (Quinlan and Hall 2010) was used with the parameter "*-bed*" to hard mask all previously detected repeats in the genome. Hard masking describes the procedure of exchanging each position covered by the coordinates inside the BED input file into ambiguous characters (N). Consequently, the genomes used for the pairwise alignments only contained non-repetitive sequences. The chromosomes/contigs of all genomes were sorted according to their length in descending order using a custom python script ("*Sort_genomes.py*"). Subsequently, a total of

42 pairwise genome alignments were generated using the aligner Nucmer implemented in the software Mummer 3.23 (Kurtz et al. 2004). Non-unique matches were excluded from the alignment ("*-mum*"). After successful alignment, any potentially remaining repetitive sequences were removed using delta-filter and the "*-g*" parameter. Finally, the coordinates of all aligned sequences were retrieved from the alignment as blocks using show-coords and the parameters "*-l*", "*-c*" and "*-d*" (Kurtz et al. 2004).

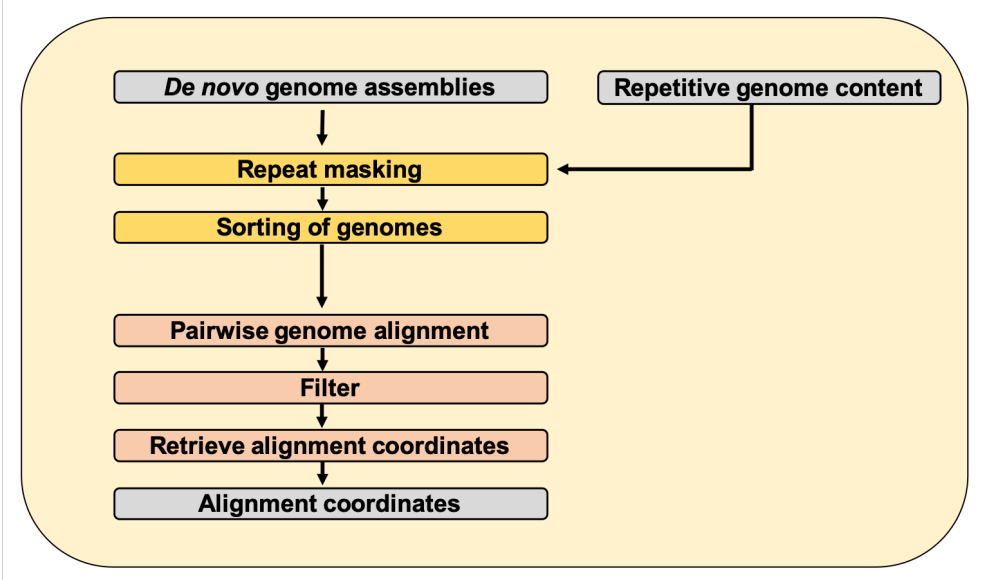


Figure 11 - Pipeline for generating a data set for the detection of intra-genus synteny. The pipeline generated a data set to study the synteny in the *Zymoseptoria* species complex. For this purpose, repeat masked genomes were aligned pairwise, and the coordinates of the generated alignments were combined into one data set per species. Grey boxes indicate input data and output data, whereas differently colored boxes indicate analyses and other processing. Yellow = Genome preparation. Red = Pairwise genome alignment and processing.

3.3.3.2. Graphical illustration of synteny in the *Zymoseptoria* species complex

For the evaluation of synteny in the genus *Zymoseptoria,* the dataset gained from pairwise genome alignments was graphically illustrated using circular and dot plots (Figure 12).

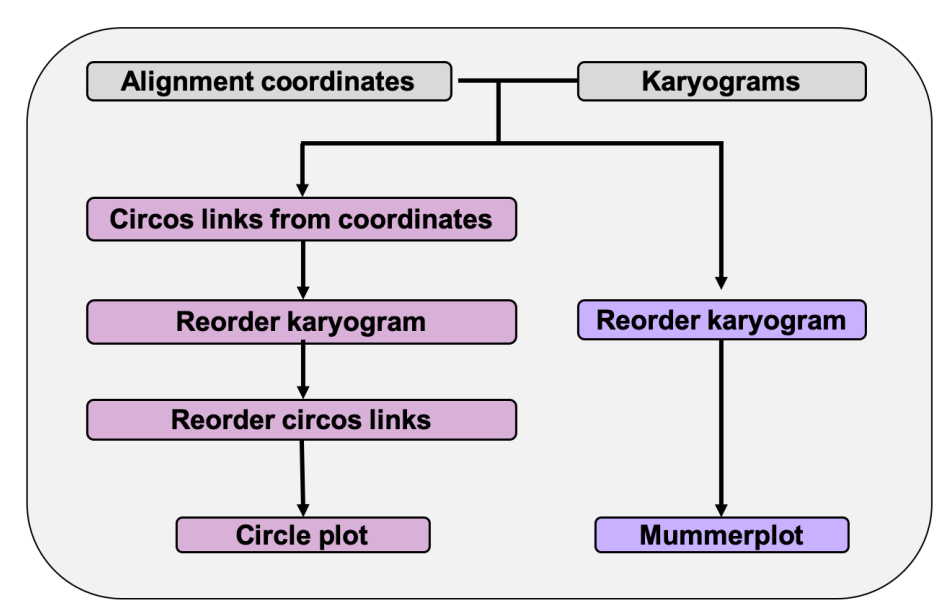


Figure 12 - Pipeline for generating graphical visualization of synteny in the genus *Zymoseptoria*. The pipeline represents the procedure for graphical evaluation of the synteny using circular (lilac) and dot plots (violet) of the pairwise genome alignments. Grey boxes indicate input data and output data, whereas differently colored boxes indicate analyses and other processing.

Visualization of macrosynteny with circle plots

The circular synteny plots were generated using the illustration software Circos (Krzywinski et al. 2009) as already described previously. The coordinate information of all alignment blocks between both reference and query genome was extracted from the pairwise genome alignments (3.3.3.1) and exported into a plain-text file using a custom python script ("*Delta2Circos.py*"). More precisely, the color and order of the connecting lines ("*SetLinkColors.py*", "*Order_Circos_Links.py*"), as well as the order of the karyograms ("*Order_Karyograms.py*"), were adjusted using custom python scripts. Finally, circular ideograms of the pairwise genome alignments were generated using Circos ("*RunCircos_Syntenic_Links.py*"). Although circular ideograms created by Circos were appropriate for evaluating the overall synteny between to genomes, they did not show smaller details like small rearrangements. Therefore, another representation strategy shown here used dot plots.

Visualization of macrosynteny with dot plots

Dot plots of all pairwise aligned genomes were generated using Mummerplot a software included in Mummer 3.23 (Kurtz et al. 2004). Both the x- and y-axes of the plots were re-sorted with the "*-R*" and "*-Q*" parameters. The files that were used for the re-sorting were generated by ordering the karyogram files that were previously used in Circos ("*Order_Karyograms.py*"). Also, chromosome-wise dot-plots were generated using the "*-r*" parameter of Mummerplot. Finally, the final pictures were exported in PNG format ("*-png*").

3.3.3.3. Detection of synteny blocks

The graphical illustrations did not allow to quantify the exact number of distinct synteny blocks, so further analyses provided this quantification. The synteny blocks were detected by the adaptation of standard methods for the detection of synteny blocks based on the genomic order of orthologous genes and applying them on a data set that harbors alignment blocks gained from pairwise genome alignments (Figure 13).

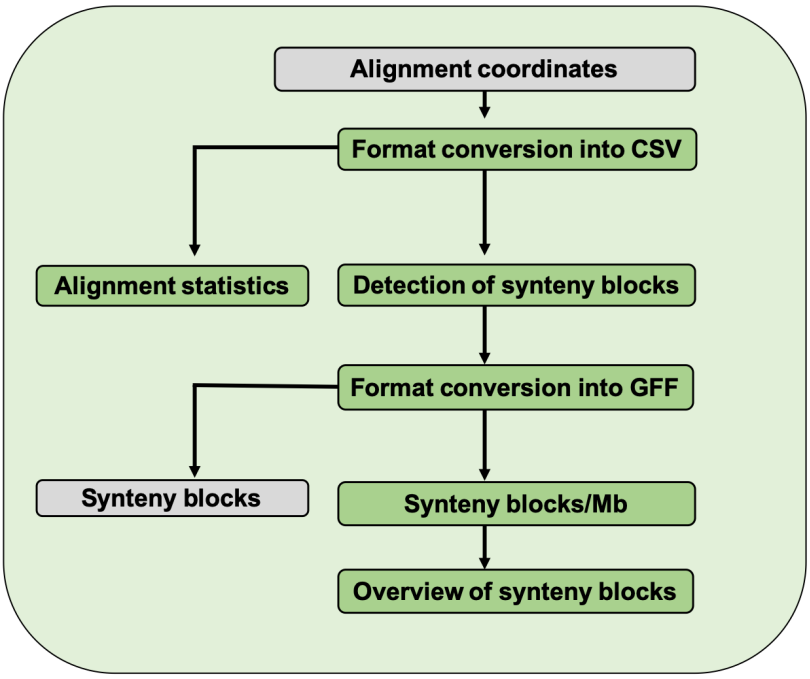


Figure 13 - Pipeline used for the detection of synteny blocks in pairwise genome alignments of seven *Zymoseptoria* species. The pipeline used the coordinates of the pairwise genome alignments to form synteny blocks. A method was adapted that was initially developed for the detection of synteny blocks from genes. To evaluate the pairwise genome alignments, alignment statistics such as the percentage of the reference genome covered by the alignment were calculated. Furthermore, the number of synteny blocks per million base pairs was determined in order to be able to compare the number of synteny blocks between the alignments. Grey boxes indicate input data and output data, whereas differently colored boxes indicate analyses and other processing.

As a first step, the alignment coordinates were converted into comma-separated value (CSV) format ("*Mummer2CSV.py*"). Further, general alignment statistics such as the average sequence identity, average block length and total length per chromosome/contig were calculated ("*AlignmentStatistics.py*"). The software DAGchainer (Haas et al. 2004), a program initially developed for orthologous genes, was used to detect synteny blocks in the dataset. By default, DAGchainer builds chains of syntenic genes in whole genomes. All alignment blocks were converted into a format compatible with DAGchainer and screened for synteny blocks ("*RunDAGchainer.py*"). The synteny blocks were subsequently converted into GFF format ("*Blocks2GFF.py*"). For each combination of genomes, two GFF files were generated. This was necessary because in later analyses each genome served as a reference. Finally, the genomic density measured as the number of synteny blocks per million base-pairs as well as the average synteny block length was calculated ("*Calc_SynB_No.py*"). The differences in the number of synteny blocks were evaluated in R ("*Synteny_Block_Evaluation.R*").

3.3.3.4. Detection of synteny breakpoints and correlation with repetitive sequences

This section aimed to answer the question of what drives the emergence of structural variation in the *Zymoseptoria* species complex. For this, the locations of the synteny blocks (3.3.3.3) were tested for correlation with repetitive element density (Figure 14).

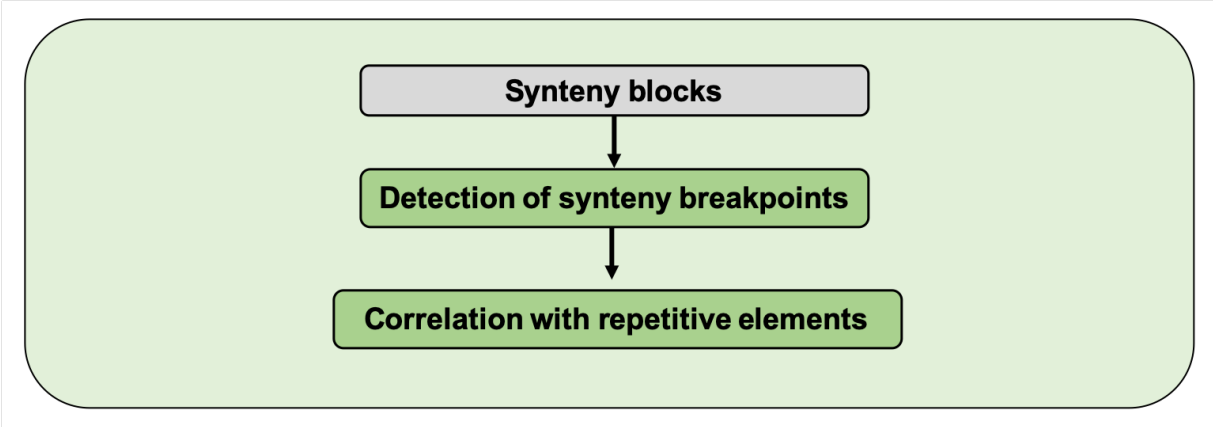


Figure 14 - Pipeline for the detection of synteny breakpoints and subsequent correlation with the distribution of repetitive genome content. The start and stop positions of the synteny blocks were defined as synteny breakpoints. Synteny breakpoints were detected in the pairwise comparisons and then examined for correlation with the positions of repetitive elements. For this purpose, a permutation test was performed. The grey boxes indicate input data and output data, whereas differently colored boxes indicate analyses and other processing.

The first step was the detection of the shared synteny breakpoints between the tested *Zymoseptoria* species ("*BreakpointDetection.py*"). A synteny breakpoint was defined as the start or stop positions of a synteny block. The synteny breakpoints were determined for all pairwise genome alignments separately. Finally, flanking regions of 5000bp were added to each breakpoint. Subsequently, all breakpoint regions were tested for correlation with repetitive element density by permutation testing using RegioneR (Gel et al. 2016) ("*Permutations_with_RegioneR_Pairwise.R*"). For testing for correlation, first, the number of overlaps between repetitive elements and synteny breakpoints was counted. In order to determine if this corresponds to a significant correlation, random loci were generated whose length was in line with those of the synteny break regions. The number of overlaps was also counted for these regions. This step was repeated a total of 1000 times and served as the basis for determining the correlation. The randomized regions of the pairwise comparisons were created only within the aligned region, while the shared synteny breaks were distributed across the entire reference genome.

3.3.4. Detection of structural variation

This study analyzed structural variation in several sets of isolates of *Z. ardabiliae* (17 isolates), *Z. brevis* (9 isolates), *Z. pseudotritici* (27 isolates) and *Z. tritici* (13 isolates) (Figure 15) using short read alignments ("*StructuralVariation_prepDataset_HYDRA.py*"). Short read mapping is a common approach for detecting structural variation because discordant read pairs, so read pairs that show abnormal alignment to the reference can be used to detect structural variation like insertions, deletions, inversions or duplications. Insertions and deletions can be, for example, detected by a larger or smaller insert size between the mapped paired reads, whereas the indication of inversions and duplications are unexpected alignment orientation (Arthur et al. 2017). For the preparation of the short read alignment, the reads were filtered using Trimmomatic (Bolger et al. 2014) as described in 3.3.2. The short read aligner BWA (Li and Durbin 2009) was used to align the short read sequencing data of 16 *Z. ardabiliae*, 8 *Z. brevis*, 26 *Z. pseudotritici,* and 12 *Z. tritici* isolates to the respective genome reference. Discordantly mapping read pairs were extracted using Samtools (H. Li et al. 2009; Li 2011). For removing false positives, the more sensitive mapping tool Novoalign (www.novocraft.com) was used to remap the extracted reads to the reference genome using sensitive settings. After the re-alignment procedure, the discordantly paired

reads were extracted and processed using Samtools. A detailed description of the dataset preparation can be found in the supplement (8.7). Finally, Hydra-Multi (Lindberg et al. 2015) an extension of the software Hydra (Quinlan et al. 2010) used the discordant reads for the detection of structural variants (Figure 15 & Figure 16).

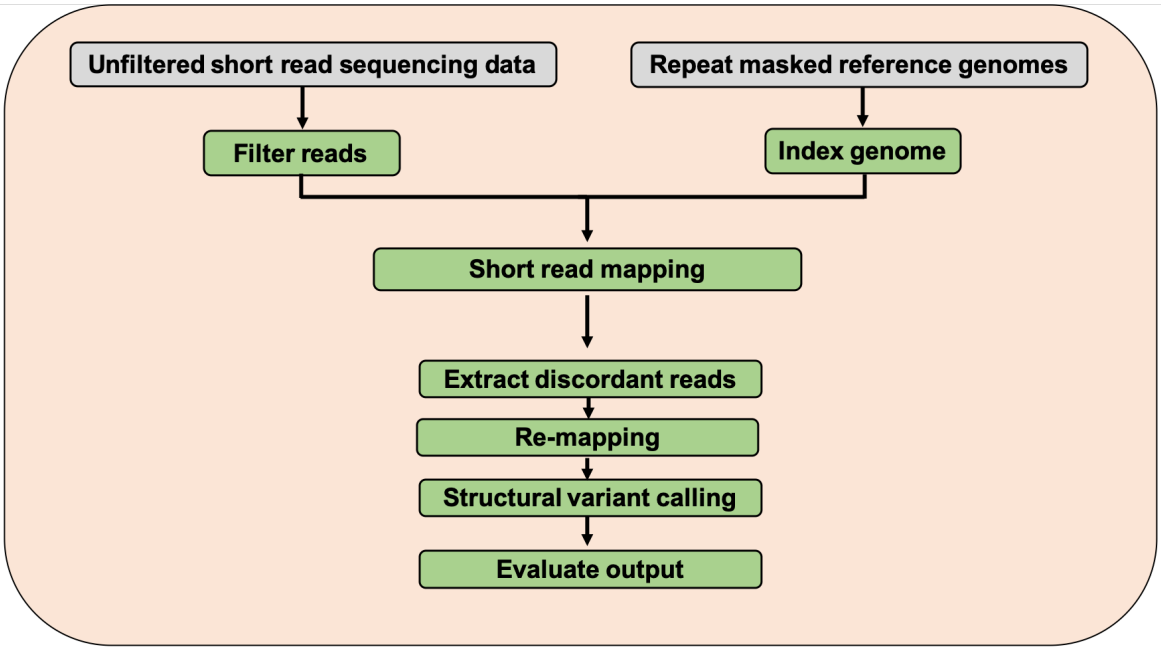


Figure 15 - Pipeline for the detection of structural variation in the genome datasets of *Z. ardabiliae*, *Z.brevis*, *Z. pseudotritici,* and *Z. tritici*. To detect the structural variation a method based on short read sequences was used. For this purpose, first short reads were aligned against a repeat-masked reference genome. Afterward, the discordantly aligned reads were extracted from the alignment and re-aligned using a more sensitive aligner. The discordantly aligned reads were again extracted from the alignment and used for the detection of the structural variations. Grey boxes indicate input data and output data, whereas differently colored boxes indicate analyses and other processing.

As a first step of the hydra pipeline an input file for Hydra-Multi that defined the name of each mapped genome dataset and the exact path to the alignment file was generated ("*make_hydra_config.py*"). Next, the final dataset of discordant reads was extracted using "*extract_all_discordants.sh*" while operating on 16 processors. Remaining duplicated read pairs were removed using the python script "*dedupDiscordantsMultiPass.py* with the "*-s*" parameter set to 3. This parameter sets the number of base pairs needed to overlap to mark a duplicate. Subsequently, all alignments on the same chromosome and the same orientation were routed into the same file as preparation of assembly ("*hydra-router*"). Next, all previously prepared chromosome/orientation subsets were assembled into breakpoint contigs (breaktigs) using 16 processes and the maximum allowed read depth set to 150 ("*assemble-routed-files.sh*"). Finally, all

assembled subsets were combined ("*combine-assembled-files.sh*") in a single file, and the final breakpoints were called based on the breaktigs ("*finalizeBreakpoints.py*"). A custom R script was used to evaluate the results ("*Eval_Hydra_out.py*").

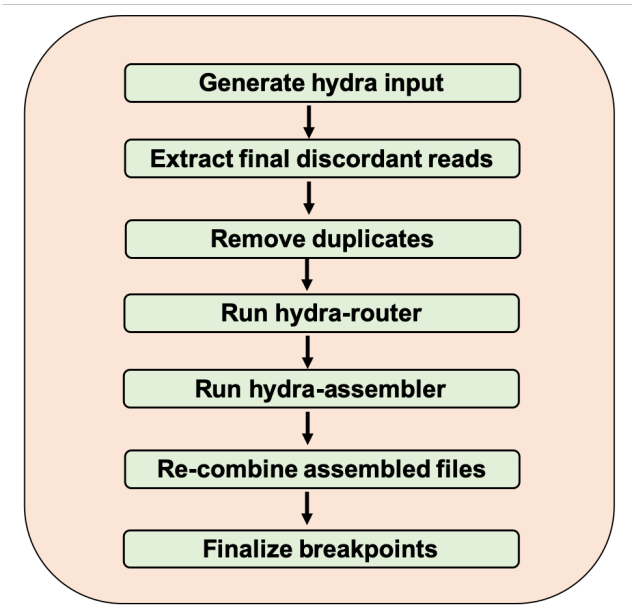


Figure 16 - Procedure for the detection of structural variants implemented in the software Hydra-Multi. The analysis of discordantly mapped reads on structural variations was performed using the Hydra-Multi software. The pipeline used by this software consisted of processing the discordantly aligned reads into the compatible input format. After extracting the discordantly mapped reads from the dataset and eliminating duplicated reads, the readings of a chromosome were translated into their file based on their orientation. The resulting sub-sets were subsequently assembled into so-called breakpoint contigs (breaktigs) and merged into a single file. The final detection of the structural variation was performed based on the breaktigs**.** The figure shows the pipeline of Hydra-Multi (Lindberg et al. 2015)

3.3.5. Generating multiple genome alignments

This study analyzed variation in and between species of the *Zymoseptoria* species complex. For this purpose, MGA was prepared using TBA (Blanchette 2004). The MGA was screened for genome-wide intra- and inter-specific sequence polymorphisms (Figure 17).

As a first part of the pipeline the *de novo* genome assemblies were pre-processed (Step 1) and a phylogenetic tree was prepared (Step 2). For the preparation of multiple genome alignments with TBA, all genomes were filtered from short chromosomes/contigs shorter than 1Kb ("*AssemblyFilter.py*"), and the chromosome/contig names were reformatted ("*formatSeq_PacBio.py*", "*formatSeq_SOAPdenovo.py*"). These steps were done to reduce the total runtime of the TBA and the compatibility with LastZ (Harris 2007). In a second preparatory procedure, a phylogenetic tree was generated from

Beta-Tubulin. The *de novo* genome assemblies were searched for homologs of the *Beta-Tubulin* gene of *Z. tritici* (Accession number: AJ310917) using a custom python script ("*blast_script.py*"). As a requirement of the blast analysis, the genomes were first converted into a database using the tool FormatDB (https://packages.ubuntu.com/trusty/blast2). Next, the full genome databases were searched for the *Beta-Tubulin* sequence using Blastall (https://packages.ubuntu.com/trusty/blast2). The resulting hits for each genome were collected and saved in a single file in FASTA format ("*Create_seaview_input.py*"). Subsequently, the sequences were aligned by muscle using the software Seaview 4.0 (Gouy et al. 2010). Also, with this software, an unrooted phylogenetic tree in NEWICK format was generated using a parsimony method implemented.

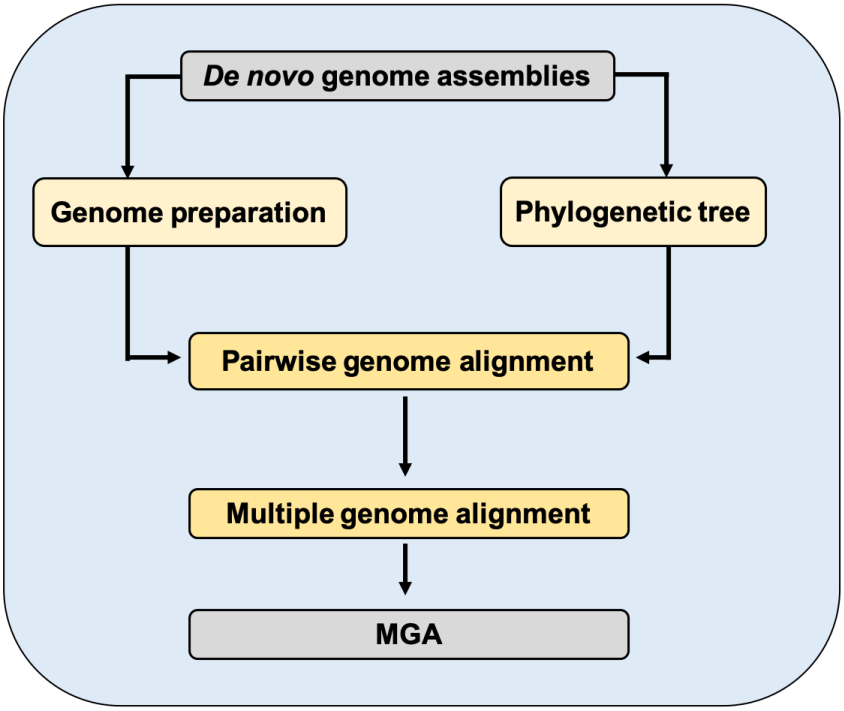


Figure 17 - Pipeline for the generation of multiple genome alignments using TBA. For preparing the multiple genome alignments, the genomes were filtered from sequences longer than 1 kb, and a single gene phylogeny based on the gene sequence of beta-tubulin was created. These steps were necessary to reduce the runtime of the alignment process as well as to create a guideline for the alignment process through the phylogeny. Based on the phylogenetic tree, pairwise genome alignments were first generated, which were then used as the basis for creating the multiple genome alignment. Grey boxes indicate input data and output data, whereas differently colored boxes indicate analyses and other processing.

In the next step, pairwise genome alignments were generated using LastZ (Harris 2007). For the preparation of the alignment procedure, a list of all possible combinations of pairwise alignments was generated using the tool all_bz and the one-gene

phylogeny as the input. With this list, the main shell script for the execution of LastZ was generated ("*all_bzToSGE_modified.py*"). Subsequently, all pairwise genome alignments were used to generate the final MGAs using Tba (Blanchette 2004). The resulting MGA was generated in the multiple alignment format (MAF) which consisted of a large number of alignment blocks (Figure 18). Each block in the MGA had a total sequence length (block length) and harbored a distinct number of sequences (block size) (Figure 18). The finished MGAs were projected against a chosen reference genome using the tool maf_project distributed with TBA. The projecting removed every alignment block in a given MGA that did not include the reference genome. Further, the projection reoriented all remaining bocks according to the reference sequence, so that the reference sequence always represented the positive strand. A custom python script calculated quality statistics of the MGAs by a custom python script ("*QMS.py*").

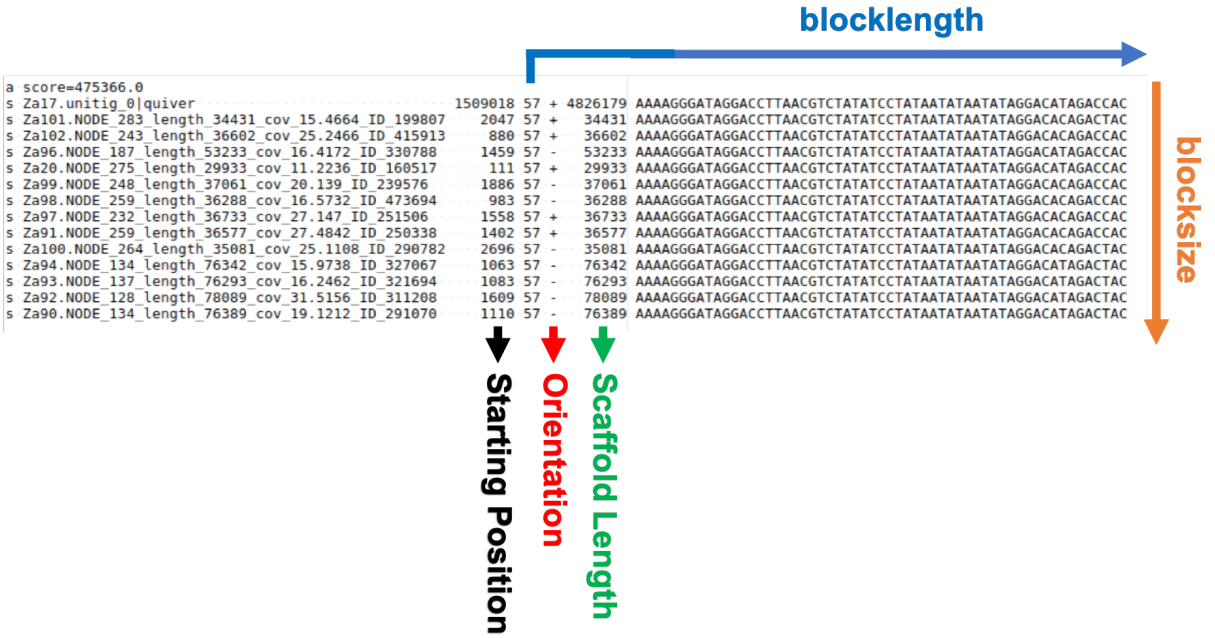


Figure 18 - Example of an alignment block in multiple alignment (MAF) format. Example picked from the projected version of the intra-species MGA of *Z. ardabiliae*. The reference sequence is Za17. A multiple alignment block in MAF format had several sequences (block size) and a length (block length). This length corresponded to the length of the alignment and not necessarily the length of the individual sequences since these could also have gaps due to the alignment. Each sequence had a starting position, orientation, and information about the total length of the scaffold the sequence is located.

3.3.6. Intra-specific diversity statistics and nucleotide variation

To assess differences of genetic diversity in the *Zymoseptoria* species complex the distribution of genetic polymorphisms were calculated genome-wide. Therefore, using the intra-specific MGAs of *Z. ardabiliae*, *Z. brevis* and *Z. tritici*, the genome-wide diversity statistics and the number of SNPs a were calculated in a window-wise (500bp windows) analysis (Figure 19).

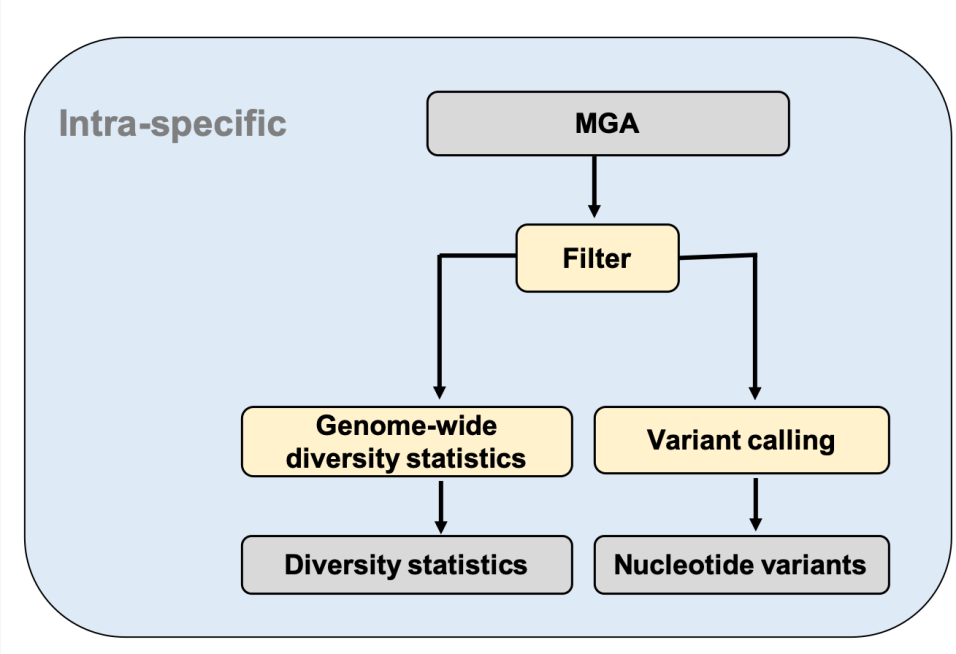


Figure 19 - Pipeline for the calculation of genome-wide intra-specific nucleotide diversity statistics and nucleotide Variation. The projected multiple genomic alignments were first filtered to regions with a high gap proportion and a block size smaller than to remove the maximum possible. Based on the filtered multiple genome alignments (MGA), diversity statistics such as estimators of nucleotide diversity (θ and π) and the derived neutrality test Tajima's D were calculated. Furthermore, polymorphic sites were detected and stored in the variant calling format (VCF) to allow further analysis. Grey boxes indicate input data and output data, whereas differently colored boxes indicate analyses and other processing.

All projected intra-species MGAs were filtered according to their block size and gap-ratio using the software Maffilter ("*Filter_Alignment_Za.bpp*", "*Filter_Alignment_Zb.bpp*", "*Filter_Alignment_Zt.bpp*"). With this filtering step, the number of false positive SNPs included in the final dataset was reduced. The software Maffilter was controlled by a plain text file (option file) with the file extension "bpp" that specified all analyses, parameter settings, and the input MGA. After filtering alignment blocks with the maximum block size and a gap-ratio of at most 64.7% in *Z. ardabiliae*, 55.56% in *Z. brevis* and 61.5% in *Z. tritici* were kept in the alignment. The filtered MGAs were used as the input for variant calling using Maffilter with the method "*VcfOutput*". This

method reported all polymorphic sites in the dataset and saved them in variant calling format (VCF; "*Maf_SNP_CALLING_Za.bpp*", "*Maf_SNP_CALLING_Zb.bpp*", "*Maf_SNP_CALLING_Zt.bpp*"). In a final pruning step, all entries with gaps or ambiguous bases were removed from the VCF files. Further, using Maffilter estimators of the observed and expected nucleotide diversity π (Tajima 1989) and θ (Watterson 1975), as well as the neutrality estimator Tajima's D was directly calculated in sliding windows of 500bp along the MGAs ("*MafDiversityStatistics_Za.bpp*", "*MafDiversityStatistics_Zb.bpp*" and "*MafDiversityStatistics_Zt.bpp*").

3.3.7. Population structure and demography

Since the results of the genome-wide calculation of Tajima's D indicated influences of population structure or demography, further analyses identified differences in the population structure or the demographic history intra-specific MGAs.

Linkage disequilibrium

Further analyses on the filtered dataset of intra-specific nucleotide variants (3.3.6) assessed the grade of clonality based on the inference of linkage disequilibrium (LD) and the population structures in the *Zymoseptoria* species complex (Figure 20).

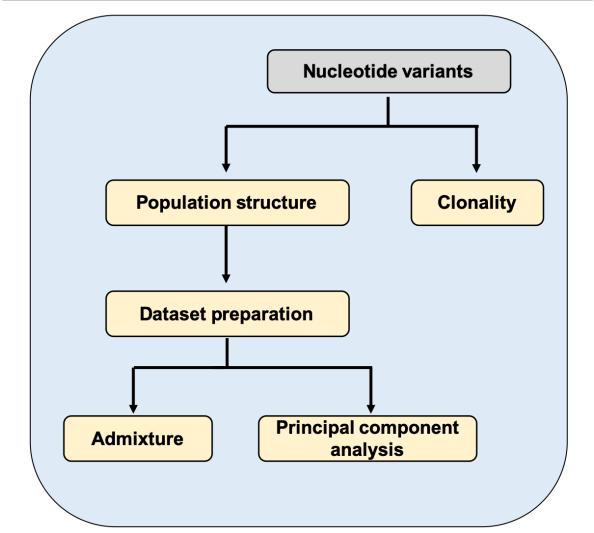


Figure 20 - Pipeline for the analysis of population structure in the *Zymoseptoria* species complex. Based on the extracted nucleotide variants, analyses of the population structure were carried out. The intra-specific nucleotide variants were examined for signs of population admixture and population structure. For this purpose, singletons and variant pairs in high LD were removed from the dataset and examined with the help of an admixture analysis and a principal component analysis. Furthermore, the dominance of asexual reproduction in the different species was investigated based on the linkage disequilibrium. Grey boxes indicate input data and output data, whereas differently colored boxes indicate analyses and other processing.

Based on the SNP dataset the amount of LD was calculated using the software package Vcftools (Danecek et al. 2011). An evaluation of the results of the complete data set in RStudio (RStudio Team 2015) was not possible because the required memory exceeded that of the system (96Gb).

Consequently, the LD analysis was performed on a reduced dataset to 40,000-43,000 SNPs ("*Run_LD_analysis.py*"). For the preparation of the analysis, all chromosome/unitig IDs and the ID of every individual in the dataset were collected by processing the nucleotide variants provided in VCF format. The whole analysis was performed for each chromosome/contig separately.

First, the number of nucleotide variants kept after reduction (t_u) was calculated as the product of the total number of nucleotide variants (t_T) per chromosome/unitig and the user-defined proportion (P_u) (**1**).

(1) $$t_u = t_T \times P_u$$

Next, the contig was divided into four equal areas for each of which the number of nucleotide variants contained was counted (t_x; x = 1 to 4). Based on t_x and t_T a weighted proportion of each of the individual parts was calculated (P_x; x = 1 to 4) (**2**).

(2) $$P_x = \frac{t_x}{t_T}$$

After the weighted proportion of variants that each region contributed to the total was known, the number of variants picked from each region (t_{px}) was calculated as the product t_u of and P_x for each part x (**3**).

(3) $$t_{px} = t_u \times P_x$$

Finally, the SNP dataset was reduced by randomly picking nucleotide variants from the parts until their number reached the respective number of variants (t_{px}). This procedure aimed to ensure that the SNP distribution in the reduced data set equals that of the complete data set so that regions with high SNP density were not sampled above average. The reduced dataset was pruned from singletons, insertions and deletions as well as SNPs with gaps or ambiguous nucleotides bad quality ("*--remove-indels*", "*--remove-filtered-all",* "*--chr*") using the functionality of the custom python script and Vcftools (Danecek et al. 2011).

After filtering linkage disequilibrium tested by calculating the correlation coefficient of the frequencies r^2 in 50kb sliding windows using Vcftools with the parameters *"--hap-r2*" and "*--ld-window-bp 50000*". This window-wise approach was performed separately for each chromosome/unitig and the results were afterwards merged using a custom

python script ("*Prepare_LD_output_Merge.py*"). The whole procedure was repeated 1,000 times, to ensure that the randomization process did not bias the results. The results were evaluated in R ("*LD_Evaluation_Summed.R*" & "*LD_Evaluation_Curve.R*").

Mating type loci

To assess the possibility of sexual reproduction, the frequency of mating type loci in the genomes of 17 *Z. ardabiliae*, 9 *Z. brevis*, and 13 *Z. tritici* isolates were tested. To determine which mating type locus was present in a genome the *de novo* assemblies were searched for homologs of the Mating type alleles *Mat1-1* and *Mat1-2* (Accession numbers: AF440399 & AF440398) of *Z. tritici* (Waalwijk et al. 2002) using the custom python script ("*blast_script.py*").

Population structure

Both the software Admixture 1.3.0 (Alexander et al. 2009) and principal component analysis (PCA) were used to analyze the intra-specific population structure (Figure 20). The pipeline started by generating subsets of nucleotide variants using a custom python script ("*SubsetVariants_Master.py*"). This script sequentially called an additional script ("*SubsetVariants.py*") that was used to exclude singletons from the analysis while keeping all SNPs in the dataset ("*-th 100%*") using the same procedure as previously explained above for the LD analysis. The filtered dataset was converted from VCF format into Plink (Purcell et al. 2007) readable format (PED). Finally, Admixture used the PED formatted nucleotide variants ("*Run_Admixture.py*") for the estimation of population sub-structure using different population numbers (k) ranging between 1 and n (n = number of individuals in the population). The program used 30 simultaneously working threads ("*-j 30*") and the "*-haploid=*"* parameter.

Differences in the cross-validation error ("*--cv 50*") were used as a main important parameter for evaluation because it allows us to identify the value of k with the best predictive accuracy in the model (http://www.genetics.ucla.edu/software/admixture/admixture-manual.pdf). The whole procedure included a series of calculations with stepwise increasing upper boundaries for k. The k value with the lowest cross-validation error in most of the previous calculation series defined the final number of populations. A custom R script generated the graphical representation of the results ("*Analyse_Admixture_Output.R*"). Additionally, a PCA represented population structure in the

Zymoseptoria species. For this, the software Plink 1.9 (Purcell et al. 2007) performed a PCA analysis on the previously defined nucleotide variants in PED format. After "*Adjust_MapFile.py*" adjusted the sequence headers inside the map file from a previous step, "*PCA_SNPs.py*" performed the actual PCA. Again, a custom R script generated the graphical visualization of the results ("*Plot_PCA.R*").

Demography

The analysis of demographic history used both the complete dataset of nucleotide variants and a subset of non-coding regions assuming that these regions are in general neutrally evolving and evolve under less selective pressure compared to the coding parts of the genome (Figure 21).

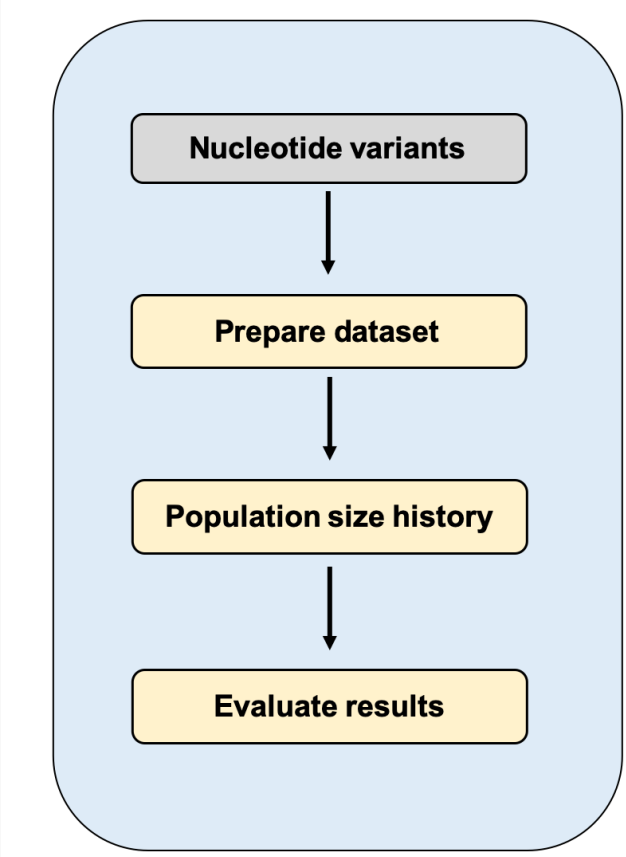


Figure 21 - Pipeline for the analysis of demographic histories in species of the *Zymoseptoria* species complex. In order to investigate differences in the demographic development between the *Zymoseptoria* species, the population size history was estimated. For this purpose, only non-coding regions were used, since it was expected that coding regions could show a biased result due to the stronger selection pressure. Thus, the generation of the final data set did not take into account the coding parts of the genome based on gene annotation. Grey boxes indicate input data and output data, whereas differently colored boxes indicate analyses and other processing.

The software uses the multiple sequentially Markovian coalescent (MSMC) model to calculate the coalescent times of local regions. It uses hidden Markov models (HMM) with hidden states consisting of the coalescent time and the labels of two lineages participating in the first coalescence event. The model makes use of the assumption that genomic regions with a high SNP density have a higher coalescence time than regions with a sparse distribution (Schiffels and Durbin 2014; Weissman and Hallatschek 2017).

Here the population size history of *Z. ardabiliae*, *Z. brevis*, and *Z. tritici* was estimated using the simulation software MSMC2 (Malaspinas et al. 2016) on four haplotypes. The full procedure was implemented in a custom python script ("*Run_MSMC.py*"). First, all chromosome/contig IDs and the ID of each were collected from the VCF file. Next, by using the intra-specific MGA, a masking file with information about the position of all callable sites was generated separately for each chromosome/contig. Additionally, since the analyses were restricted to non-coding regions, a second masking file providing the position of coding sites was prepared. The content of the second file was extracted directly from the provided gene annotation. In all methods that involve coding sequences a published gene annotation of Zt09 (Grandaubert et al. 2015) and *de novo* annotations of Za17 (*Z. ardabiliae*) and Zb87 (*Z. brevis*) kindly provided by co-workers (Sanchez-Ramirez unpublished) were used.

To prepare the creation of the MSMC2 input the SNP dataset was divided into individual VCF files, each comprising only one strain ("*--indv*") and one chromosome/contig ("*--chr*") using Vcftools. The resulting files were then also filtered using Vcftools, leaving no entries with a different filter flag as PASS ("*--remove-filtered-all*").

In the final step, all generated individual files were converted back to a single file in the MSMC2 input format. For this purpose, the developers of MSMC2 offer a useful tool on their website (https://github.com/stschiff/msmc-tools.git). However, since the provided script ("*generate_multihetsep.py* ") is not suitable for the use of haploid genomes, it could not be used in the original version.

Consequently, a modified version was developed that allowed the use of haploid datasets ("*generate_multihepsep_hapl.py*"). After the successful creation of the final data set, the MSMC2 analysis was performed using a time segment pattern of "'25*2+10*2+1*2". The remaining parameters were kept in the default.

3.3.8. Adaptive evolution in wild and domesticated environments

Positive diversifying selection

Fungal plant pathogens and their hosts may co-evolve under evolutionary trench-warfare or arms-race which are typically characterized by balancing selection or selective sweeps (Tellier et al. 2014). This thesis used the parameter ω that provides a direct measure for selective pressures on codons to detect positive diversifying selection. The parameter ω is defined as the ratio of non-synonymous and synonymous substitutions (dN/dS) (Yang and Nielsen 2002; Yang et al. 2005). Since the dN/dS can also be used for population data (Anisimova et al. 2001), this thesis applied it on an intra-specific dataset. Although dN/dS-based methods were initially developed for different species, the results of a simulation study show that divergent populations can also be studied (Kimura 1977; Goldman and Yangf 1994; Muse and Gaut 1994; Nielsen and Yang 1998). Because in essential genes only a few sites can mutate non-synonymously without negative fitness effects (Nielsen and Yang 1998), in this Ph.D. thesis models for the detection of positive diversifying selection of single codons. Methods based on the maximum likelihood approach allow the comparison of models with and without positive selection by performing a likelihood ratio test (LRT) (Goldman and Yang 1994; Muse and Gaut 1994). In this context, genes with sites under positive diversifying selection are expected to show the best support with a higher probability with models that allowed positive selection. This thesis used the software PAML for the determination of positive selection using ω under various models (Yang 1997; Yang 2007).

Further, the McDonald-Kreitman (MK) test was used as a measure of adaptive evolution (McDonald and Kreitman 1991). For the detection of genes that may have undergone adaptive evolution, a complex pipeline that used intra-specific MGAs was used (3.3.5) (Figure 22). Based on the number of cores used the annotations described in 3.3.7 were modified and divided into several smaller subsets using the scripts "*Modify_Annotation_CDS_IDs.py*" and "*Divide_Gene_List.py*". In the next step, a python script was created for each annotation subset that performs the entire search for signatures of positive diversifying selection for the particular subset ("*Create_Python_Scripts.py*"). The python scripts were created by including the path to the annotation file subset (X) to two prepared script templates. ("*get_Evolutionary_rate_template_12*" & "*get_Evolutionary_rate_template_78*"). Finally, a master shell script was generated includes the list of commands that execute the previously generated python

scripts ("*Create_Master_Shell_Script.py*"). The scripts that were called by the master shell script started the whole detection pipeline of positive selection on a single gene annotation subset (black dashed box, Figure 22). By running one of the python scripts ("*get_evolutionary_rate_PartX.py*"), a gene-wise screening for genes with signatures of positive diversifying selection was started. Because the analysis tested only coding sequences, all CDS for each gene were extracted from the MGA (Dutheil et al. 2014)

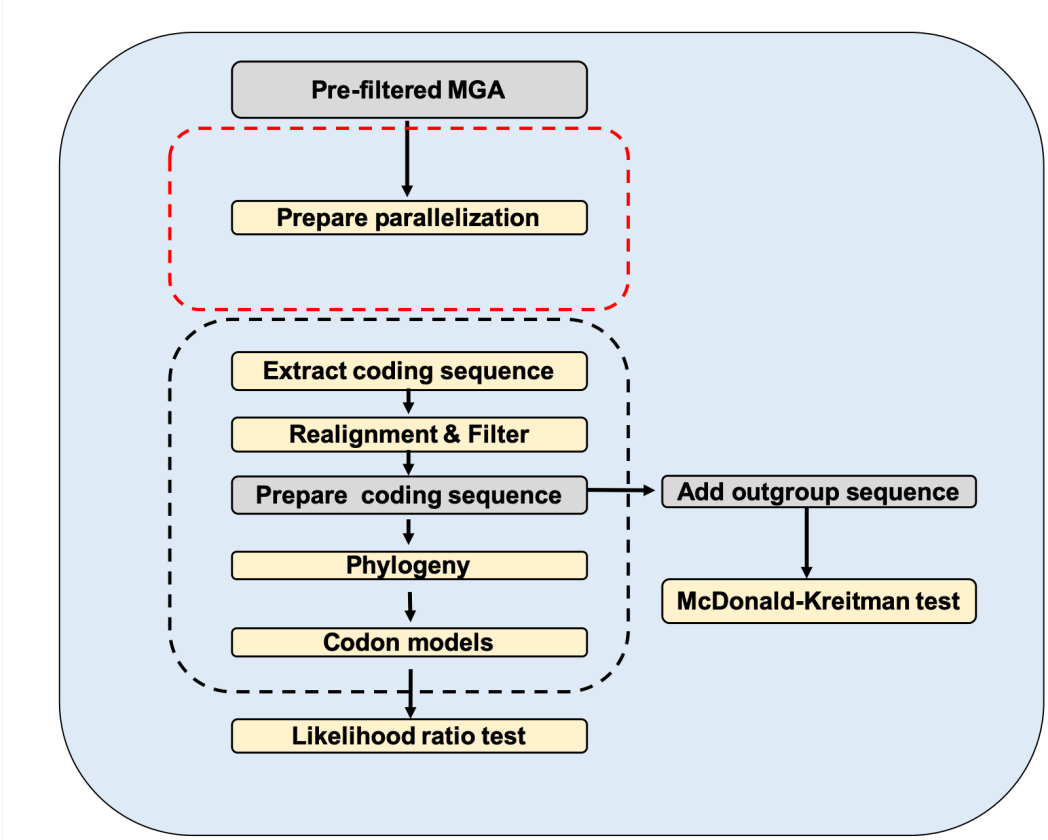


Figure 22 - Screening for genes that may have undergone adaptive evolution. The pipeline used two measures based on the number of non-synonymous and synonymous substitutions and polymorphisms to find genes under positive selection. For the detection of genes with signs of positive selection, a pipeline was developed that sought signatures on positive diversifying and positive directional selection. Thus, the coding sequences of each gene were extracted from the multiple genome alignment, assembled and filtered. If necessary, the frameshift has been corrected. For the detection of genes with signatures of diversifying selection, individual codons exhibiting positive selection in the gene were detected by comparison of different statistical models. In each case, two models were compared with each other differed in whether positive selection was allowed or not. A subsequent likelihood ratio test was used to evaluate the results. However, by adding the homologous sequences from an outgroup species (*Z. passerinii*, Zpa63) to the ultimate genome alignment, the data set was further prepared for analysis with the McDonald Kreitman test (MK test). The significance of the individual results of the MK test was determined with Fisher's exact test. The p-values of both analyzes were corrected using the false discovery rate. Grey boxes indicate input data and output data, whereas differently colored boxes indicate analyses and other processing.

After concatenation of the extracted CDS, the gene sequence was realigned with Mafft (Katoh et al. 2009) and filtered according to the number of gaps (max. 5%). Alignments with less than three sequences were excluded from the analysis. After concatenating all CDS per gene, and if necessary, Macse (Ranwez et al. 2011) was used to perform a frameshift correction. As a requirement for using Codeml, a part of PAML, a phylogenetic tree was generated for each gene the script using the software Phyml v20160207 (Guindon et al. 2010). Subsequently, several codon models to the aligned sequences of each gene using Codeml. In principle, the screen for positive selection based on the comparison of likelihoods of a model allowing positive selection (M2a and M8) and a simpler model missing it (M1a and M7). To be more precise, each gene was tested for its support of the different models using two pairs of models (M1a/M2a and M7/M8). For each gene, the best fitting model per combination was determined by an LRT ("*Statistical_Test_for_Positive_Selection.py*"). Finally, a correction for multiple testing of all p-values computed from the LRT was done using the false discovery rate (Benjamini and Hochberg 1995). The final dataset of genes with signatures of positive diversifying selection was determined by using the significance threshold of 0.01 ("*Positive_Selection_Evaluation.R*"). The final list of genes with signatures of diversifying selection, i.e., those genes that were significant in both model comparisons, was created using a custom R scripts ("*Draw_PS_plots.R*").

Detection of past positive directional selection – McDonald-Kreitman test

For the MK test, the gene alignments that were already used in the analysis of diversifying selection were added by one sequence of *Z. passerinii* ("*Add_Outgroup.py*"). The genome sequence of the *Z. passerinii* isolate Zpa63 was searched for the particular gene of interest using a blast search ("*blast_script.py*"). The number of blast hits was reduced by filtering of the hits using an E-value threshold of 1e-05. After removing possible overlaps of the single hits, the hits were placed in the correct order and, if necessary, reverse complemented. After merging the filtered hits into a single sequence and as long as the sequence covered at least 90% of the query, the sequence was added to the multiple sequence alignment with Mafft. After frameshift correction by Macse, the nucleotide variants were detected using Maffilter.

Further, the gene coordinates were extracted and modified directly from the gene annotation and saved the result as GFF using a custom python script "*GenGFF_Files.py*". Subsequently, the R package Popgenome (Pfeifer et al. 2014) was used to perform an MK test using the nucleotide variants, the multiple sequence alignment and the GFF

file as inputs. Filtering of the results for significance was done using a Fisher's Exact Test with a threshold of 0.05 ("*Positive_Selection_Evaluation.R*").
To further narrow down potential candidates with a role in virulence, the resulting gene lists of the genes with signatures of positively diversifying or positive directional selection were compared using "*Draw_PS_plots.R*". Further, the software Blast2Go (Götz et al. 2008) was used to screen candidate genes for homologies with known proteins of other pathogens. The analysis was performed using a blast analysis with "fast-blastp", a gene ontology (GO) mapping, and a final annotation, with default settings. From the resulting gene list genes were further analyzed by first searching for protein domains using the web tool Smart (Letunic et al. 2015; Letunic and Bork 2018) and subsequent prediction of active sites inside of the domains using the web tool Coach (Yang et al. 2013a; Yang et al. 2013b) both with standard conditions. This step was performed to determine if the sites that were predicted by Codeml to be under positive diversifying selection are part of the active sites of the protein. Sites under positive diversifying selection at which non-synonymous substitutions occur at a higher rate than synonymous ones were determined with the Bayes Empirical Bayes (BEB) calculation of posterior probabilities for site classes (Yang et al. 2005), implemented in Codeml (PAML, models M2a and M8).

Detection of effector candidates

Effector candidates were detected using the already described gene annotations the amino acid sequences of all annotated genes in *Z. ardabiliae*, *Z. brevis* and *Z. tritici* (Figure 23).
First, the amino acid sequences were extracted from the gene annotations of *Z. ardabiliae*, *Z. brevis* and *Z. tritici* using a custom python script ("*Prepare_protein_fasta.py*"). After retrieving all sequences, a pipeline based on three consecutive steps was used to screen for effector candidates ("*Run_Effector_Screen.py*"). First, proteins with a putative secretion signal were detected using, SignalP 4.1 (Petersen et al. 2011). Next, TargetP 1.1 (Emanuelsson et al. 2000) was used to estimate the cellular localization of all proteins with a secretion signal. Finally, effector candidates among all genes with association to the secretory pathway were predicted using EffectorP 1.0 (Sperschneider et al. 2016).

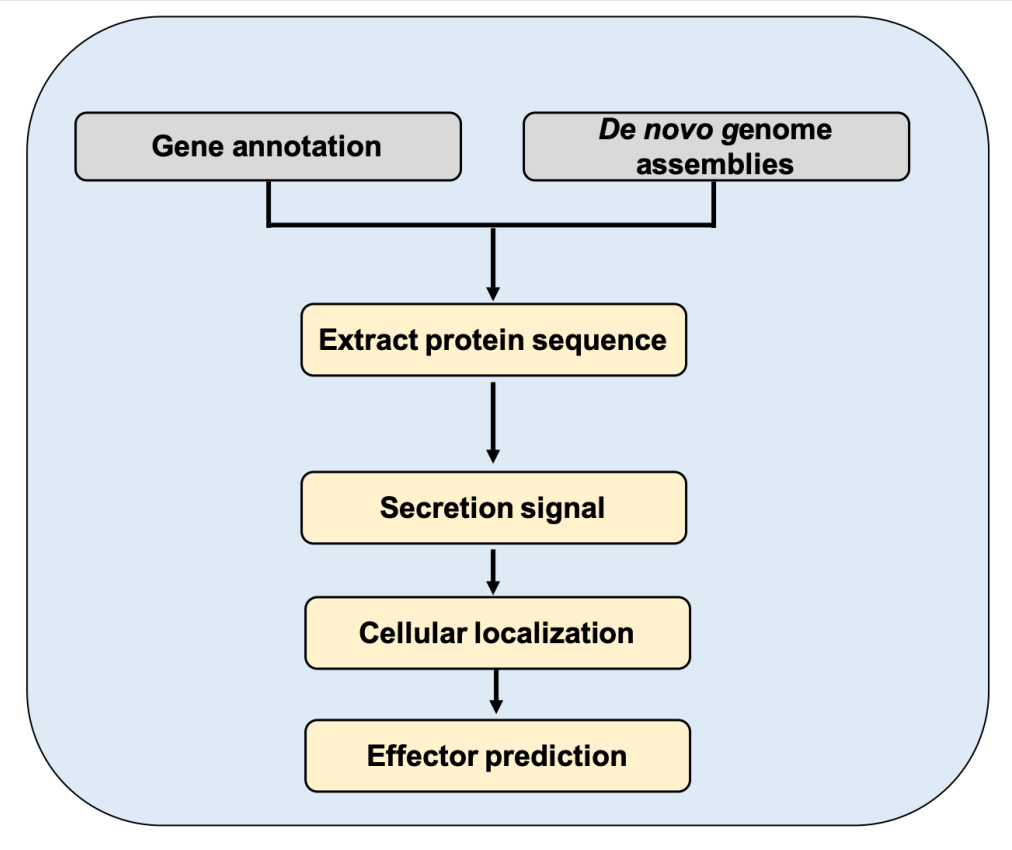


Figure 23 - Pipeline for the detection of effector candidates. To estimate effector candidates in the reference genomes of *Z. ardabiliae* (Za17), *Z. brevis* (Zb87) and *Z. tritici* (Zt09), the protein sequences of all genes were first examined for potential signal peptides. The genes tested positive for signal peptides were then screened for their intriguing cellular localization. Finally, to find effector candidates, all proteins that had a possible link to the secretory pathway were examined for typical effector properties. Grey boxes indicate input data and output data, whereas differently colored boxes indicate analyses and other processing.

Comparative analyses of the gene sets under adaptive evolution or effector candidates

The effector candidates were tested for their enrichment in genes with signatures of positive diversifying or positive directional selection ("*Compare_Lists.py*" & "*PS_Test.R*").

Further, to gain a more in-depth insight into the importance of diversifying selection or the effector candidates to host specialization the corresponding lists of effectors or genes with signatures of positively diversifying selection were compared with a list of species-specific genes. First, the online tool OrthoVenn (Wang et al. 2015) was used to determine the species-specific genes by comparing existing gene annotations of *Z. ardabiliae*, *Z. brevis,* and *Z. tritici*. Second, the lists of species-specific genes were compared with those of the effector candidates and positive diversifying genes ("*Compare_Lists.py*"). Finally, over- or under-representation of effector candidates or positive diversifying genes among species-specific genes was tested ("*PS_Test.R*").

3.3.9. Phylogeny of the *Zymoseptoria* species

3.3.9.1. Phylogenomics from multiple genome alignments

Computing inter-specific phylogenies allowed to determine the phylogenetic relationship of the *Zymoseptoria* species (Figure 24). An inter-specific MGA of 65 *Zymoseptoria* isolates of 5 different species projected against the *Z. tritici* reference strain Zt09 was used as an input for the analysis. Previous to the analyses, Mafft (Katoh et al. 2009) was used to realign all blocks in the multiple genome alignments locally. The software Mafft was applied to each alignment block separately using the method "*SystemCall*" of Maffilter ("*MAFFTRealign.bpp*"). Only alignment blocks with a length of 500bp or longer were kept after the realignment. Further filtering the MGA quality was improved by filtering of the re-aligned MGA for maximum block size and a maximum gap-ratio of 60% using Maffilter ("*Filter_Alignment_1.bpp*") and 50% ("*Filter_Alignment_2.bpp*").

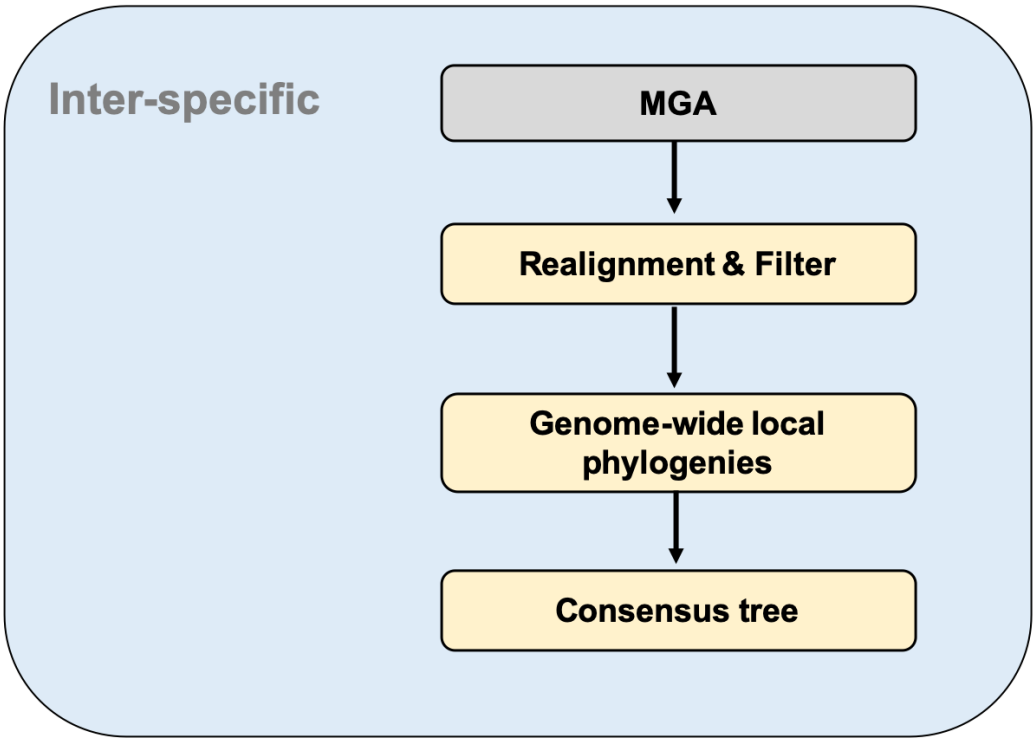


Figure 24 - Pipeline of the phylogenomics analyses for the analysis of intra-specific phylogeny. For the phylogenetic classification of the individual *Zymoseptoria* isolates in the genus, a phylogenomic tree based on genome-wide nucleotide variation was constructed. For this purpose, an inter-specific multiple genome alignment from in total 67 isolates of 5 *Zymoseptoria* species. The final phylogenomic tree was created as supertree from window-wisely computed unrooted phylogenies. Neighbor-joining was used to create the individual phylogenetic trees. Grey boxes indicate input data and output data, whereas differently colored boxes indicate analyses and other processing.

After filtering, a phylogeny based on genome-wide nucleotide substitutions was calculated using Maffilter ("*MafTreeStatistics.bpp*"). Pairwise distance matrices were calculated window wise (1Kb windows) using the method "*DistanceEstimation*". The phylogenetic tree was generated using neighbor-joining (NJ) ("*DistanceBasedPhylogeny*"). Finally, the output trees were saved in Clustal Tree format (*.dnd) for each window ("*OutputTrees*"). A custom R-script ("*S2_SUBS.R*") was used to modify the tip labels of the resulting trees. A consensus tree was generated from the modified trees using the program Consense implemented in the Phylip v3.696 package (Felsenstein 1989; Felsenstein 2005) with the majority rule and the genome of *Z. passerinii* (Zpa63) as the outgroup. The software Ugene (Okonechnikov et al. 2012) was used to visualize the resulting phylogenetic tree as a cladogram.

4. Results

4.1. Phylogenomics of the *Zymoseptoria* species complex

To generate the phylogenetic tree, a multiple genome alignment of 65 isolated from 5 *Zymoseptoria* species was generated. The raw alignment had a total length of 24.7Mb and a length of 3.32Mb after filtering (Figure S 23- 24). The phylogenomic tree was generated as a consensus-tree from 3,315 single trees constructed from genome-wide SNPs (see Materials and Methods 3.3.9) provided an overview of the phylogenetic relationships in the *Zymoseptoria* species complex (Figure 25).

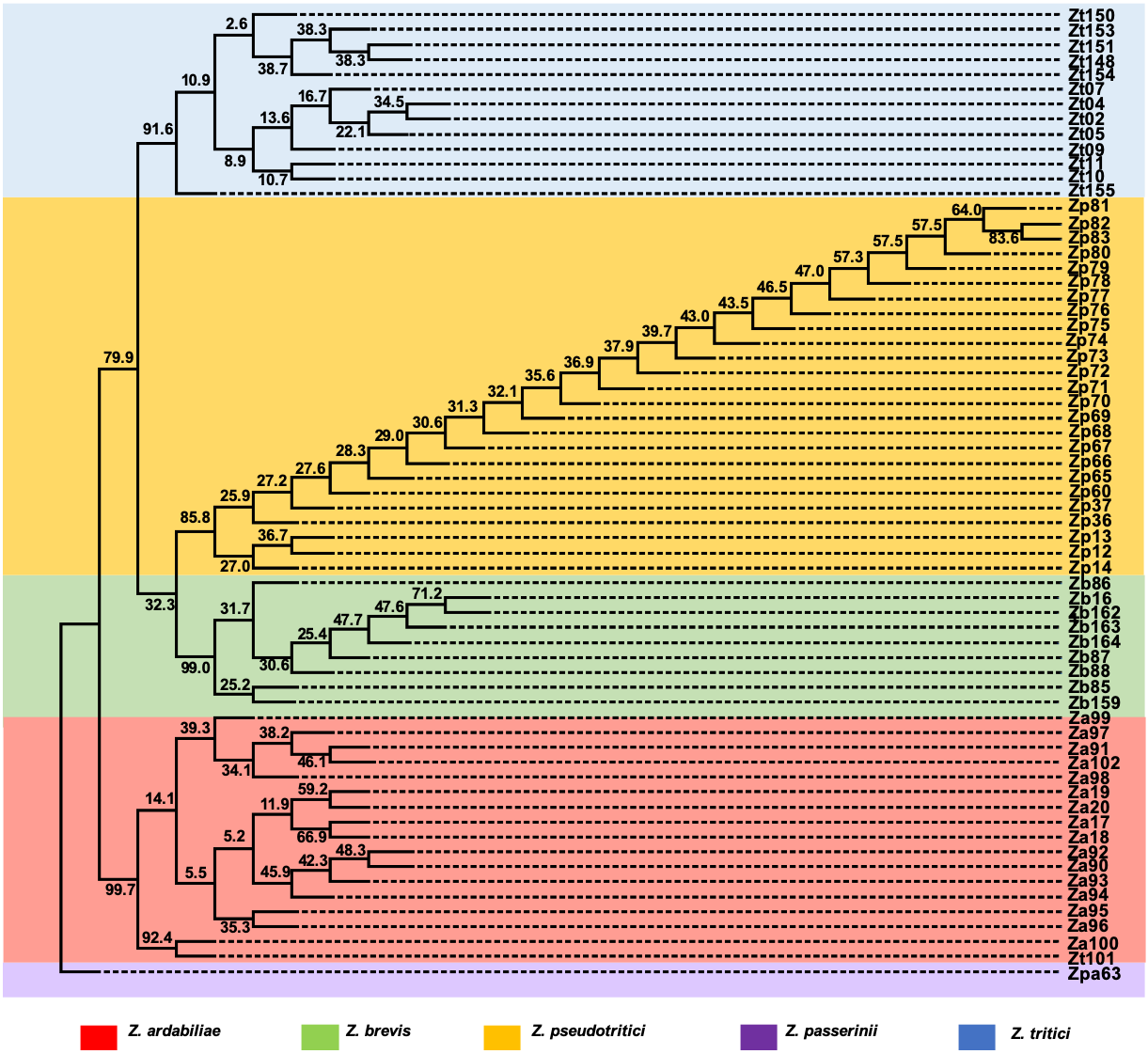


Figure 25 - Phylogenetic tree of the *Zymoseptoria* species complex based on genome-wide single nucleotide variations. The phylogenomic tree of the *Zymoseptoria* species complex showed a close relationship between the wild grass pathogenic *Z. brevis* and *Z. pseudotritici*. Furthermore, the tree showed that the wheat pathogen *Z. tritici* is more closely related to *Z. brevis* and *Z. pseudotritici* than to the barley pathogen *Z. passerinii* or the wild grass pathogen *Z. ardabiliae*. The numbers next to the branches refer to the percentage of trees that support the branch of this supertree.

The cladogram of the unrooted phylogenetic tree showed a close relationship between the wild grass associated fungi *Z. pseudotritici* and *Z. brevis*. Interestingly, these two fungi were closer related to the wheat pathogen *Z. tritici* than to the other wild grass associated pathogen *Z. ardabiliae*. However, in the phylogeny the barley pathogen *Z. passerinii* separated from the other members of the species complex and was therefore used as an outgroup for comparative methods such as the McDonald-Kreitman test (see below).

4.2. Genome plasticity and synteny of the *Zymoseptoria* species complex

In *Z. tritici* previous studies showed high genome plasticity and presence-absence polymorphisms of accessory chromosomes (Wittenberg et al. 2009; Goodwin et al. 2011). Overall the results of this Ph.D. thesis demonstrated genome plasticity characterized by the presence of accessory chromosomes as conserved traits of the genus *Zymoseptoria*, and it further showed a link between repetitive elements and synteny breaks. Further this thesis confirmed macrosynteny in the genus *Zymoseptoria*.

4.2.1. Genome plasticity is common in the *Zymoseptoria* species complex

Genome lengths and repeat content of the Zymoseptoria species differ inter- and intra-specifically

The bioinformatic analyses in this thesis used *de novo* genome assemblies of five different *Zymoseptoria* species (see Materials and Methods 3.3.2). Some assemblies were generated in cooperation with or received from collaborators (Daniel Croll, University of Neuchâtel). The *de novo* genome assemblies showed differences in the proportions of repetitive sequences as well as species-spanning genome length polymorphisms. The Illumina-based assemblies had lengths of 30.7-38.4Mb, comprising 760-2,222 scaffolds and overall N50 values of 60.3-150.0 Kb (Table 3). Detailed information about the short-read assemblies is included in the supplement (Table S 2). Assembly methods based on SMRT technology generated larger *de novo* assemblies with total lengths of 39-42Mb, divided into 23-183 contigs and N50 values of 0.5-2.93Mb (Table 4).

Table 3 - Summary statistics of filtered genome assemblies of *Z. ardabiliae*, *Z. brevis*, *Z. tritici* and *Z. pseudotritici* based on Illumina short read sequencing. The Illumina sequencing-based *de novo* assemblies showed a large intra-specific range of lengths, scaffold numbers, and N50 values. N = unknown nucleotide. # = number. The *de novo* assemblies generated here can be found on the attached USB-stick (8.15).

Species	No. Isolates	Length [Mb]	Scaffolds [#]	N50 [Kb]	$\frac{N}{100kb}$	Assembler	References
Z. ardabiliae	16	33.6–35.3	760–2,222	60.3–130.0	1.37–163.8	SPAdes	This study
Z. brevis	8	30.7–33.9	865–2,038	66.4–150.0	405.8 – 1961.5	SOAPdenovo2	This study
Z. pseudotritici	24	33.5–34.0	1,074–1,324	60.7–106.3	0 – 3,581.1	SOAPdenovo2	This study
Z. tritici	9	32.8–38.4	951–1,541	79.7–102.4	558.7–2,964.0	SOAPdenovo2	Grandaubert et al. 2017

Table 4- Summary statistics of genome assemblies of *Z. ardabiliae*, *Z. brevis, Z. passerinii, Z. pseudotritici* and *Z. tritici* based SMRT sequencing. The assembly statistics showed large assemblies, divided into low numbers of contigs with high N50 values. Repetitive element content differed on the inter- and intra-specific level. Screening for a simple telomeric repeat indicated the successful assembly of up to 19 chromosomes in Zt10. N = unknown nucleotide. * = Reference genome of IPO323 retrieved by Sanger sequencing (Goodwin et al. 2011). The *de novo* assemblies generated here can be found on the attached USB-stick (8.15).

Strain	SMRT Cells #	Coverage	Length [Mb]	Contigs #	N50 [Mb]	Repetitive [Kb]	Repetitive %	Complete Chromosomes #
Za17	5	75x [a]	38.96	59	1.41	8,369	21.48	3
Zb87	7	55x [a]	42.08	53	2.75	12,582	29.90	9
Zp13	6	73x [a]	40.22	38	2.12	9,049	22.50	5
Zpa63	1	72x [b]	41.63	183	0.50	12,866	30.91	0
Zt05	4	54x [a]	42.23	72	2.45	9,770	23.14	12
Zt10	5	69x [a]	39.43	23	2.93	7,331	18.59	19
Zt09*	---	---	39.69	21	---	7,237	18.23	20

a – Median Coverage. Calculated because a few high coverage outliers skew the distribution of the coverage values.

b – Mean Coverage. Mean coverage as calculated by SMRT link v5.1.0 (© PacBio).

However, *de novo* assembly methods based on short read sequencing technologies (i.e., Illumina sequencing), unlike those using the results of long read sequencing technology (i.e., SMRT sequencing), often have problems with more extended repetitive sequences (Phillippy 2017). Therefore, the observed differences between the assembly methods may indicate different repetitive element levels in all species tested. The detection of repetitive elements in the *de novo* assemblies constructed from SMRT sequencing showed inter-specific differences with total contents ranging between 18% in *Z. tritici* and up to 30% in *Z. brevis* (Table 4). Interestingly, even on the intra-species level, *Z. tritici* showed different repetitive element contents with values ranging between 18 and 23 percent (Table 4).

The distribution of a telomere-specific repeat in the *de novo* reference assemblies indicated that they contained up to 19 complete chromosomes (Table 4, Figure 26, Figure S 2-3). Furthermore, the results showed that genome assemblies with less repetitive sequences, such as those of *Z. tritici*, also had a higher number of complete chromosomes (15-19 entire chromosomes). The genome of the *Z. tritici* isolate IPO323 (Zt09) (Goodwin et al. 2011) served as a quality standard for the examination of telomeric repeats (Figure S 4). As in the original publication by Goodwin and co-workers (Goodwin et al. 2011), the results showed telomeric repeats at both ends of chromosomes 1-20.

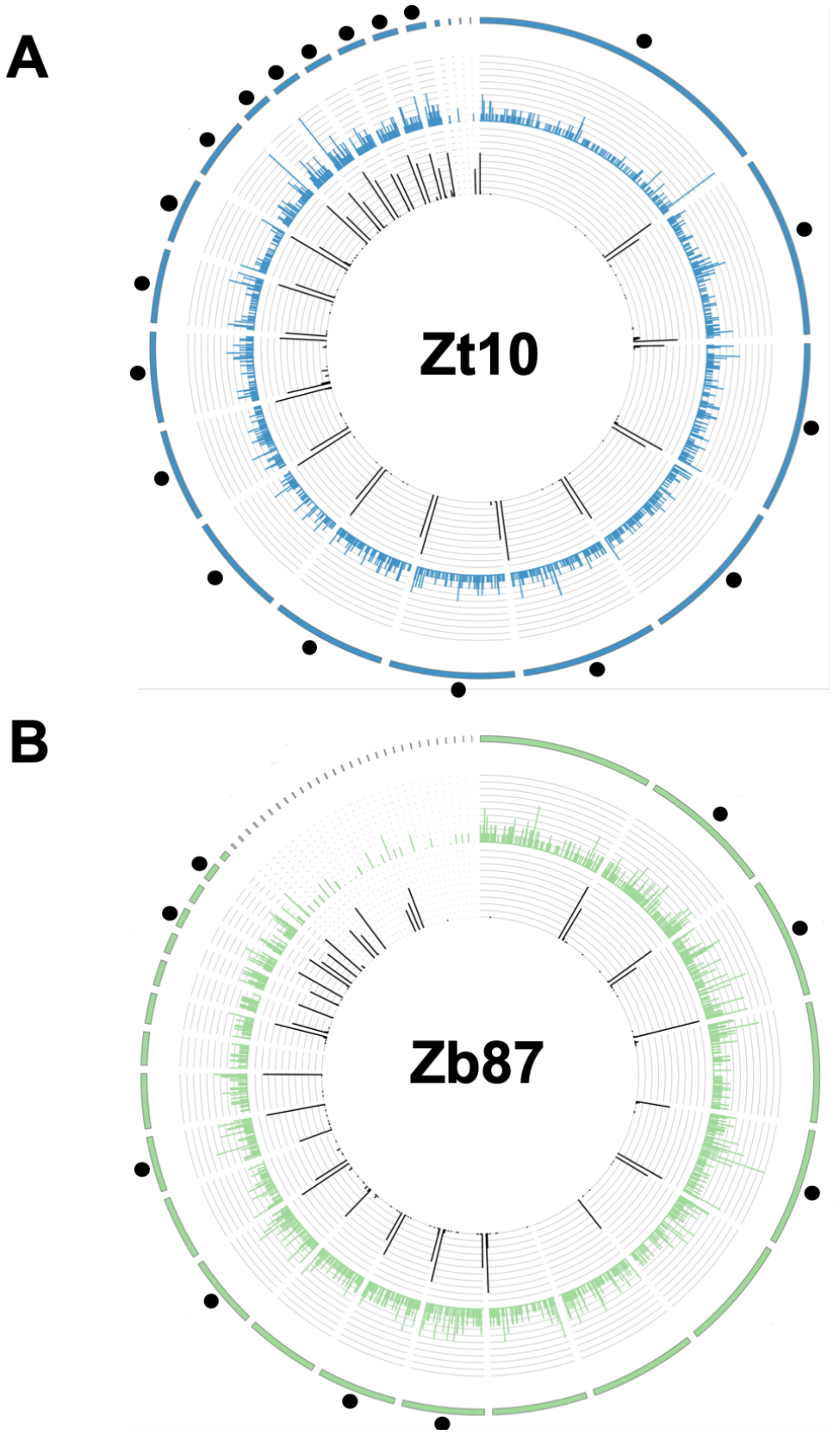


Figure 26 - *de novo* genome assemblies of Zb87 (*Z. brevis*) and Zt10 (*Z. tritici*). A) The genome assembly of Zt10 showed 19 completely assembled chromosomes (black points) and a homogenous distribution of repetitive elements. B) The genome of Zb87 showed nine entirely sequence chromosomes. The graphs show an even distribution of repetitive elements across the genome. Colored histogram track = density with repetitive sequences per window. Black histogram track = Density of telomeric repeats within a sliding window.

The presence of accessory chromosomes is conserved in the Zymoseptoria species complex

Not only insertions and deletions due to repetitive elements but also the appearance of chromosome number polymorphisms such as accessory chromosomes may be a cause of the observed intra-specific genome length polymorphisms (McDonald and Martinez 1991; Cuomo et al. 2007; Mehrabi et al. 2007; Plissonneau et al. 2016). Accessory chromosomes play a role for genome length polymorphisms (Wittenberg et al. 2009; Goodwin et al. 2011), and this Ph.D. study analyzed the closely related sister species for the presence of accessory chromosomes (see Materials and Methods 3.2.2, 3.3.3.1 and 3.3.3.2). The results showed intra-specific chromosomal length polymorphisms of the smallest chromosomes (lengths < 2.2Mb) in all wild grass-associated fungal plant pathogens (Figure 27-30 A). The strains of the wild grass associated phytopathogenic fungi showed intra-specific polymorphisms in chromosome number (*Z. ardabiliae* = approx. 4-6 chromosomes; *Z. brevis* = approx. 1-3 chromosomes; *Z. pseudotritici* = approx. 3-5) and sizes (*Z. ardabiliae* = approx. 0.225-0.85Mb, *Z. brevis* = approx. 0.45-1.4Mb, *Z. pseudotritici* = approx. 0.5-1.2Mb). The size patterns of the wild grass pathogens showed presence-absence polymorphisms of larger (up to approx. 1.2Mb) and smaller (up to approx. 0.225Mb) chromosomes than *Z. tritici* (0.4-0.8Mb) (Goodwin et al. 2011). However, the accessory chromosomes of *Z. tritici* showed homology with large numbers of wild sister species genomes (Figure 27-30 B). The homologous regions in *Z. ardabiliae* (Figure 28 B) and *Z. brevis* (Figure 28 B) shared less synteny with the accessory chromosomes of *Z. tritici* than *Z. pseudotritici* (Figure 29 B). The circular ideograms of the synteny between Zt09 (*Z. tritici*) and the other *Zymoseptoria* species showed homology of accessory sequences on the inter-species level (Figure S 14 - 19).

However, in all sister species, few of the small chromosomes were completely sequenced, and yellow stars marked chromosome bands of similar lengths on the pulsed-field gel (Figure 27-30 A). For example, the pairwise synteny analysis suggested that the accessory chromosome 16 (607,044bp) of *Z. tritici* probably was also present in *Z. pseudotritici* as it showed a high synteny with the completely sequenced unitig_32 (601,333bp) (Figure 29 B). Intra-specific comparisons using the same technique indicated, in addition to the already known presence of accessory chromosomes, the presence of an unknown accessory chromosome (unitig_15; 634,198bp) in the *Z. tritici* isolate Zt10 (Figure S 13).

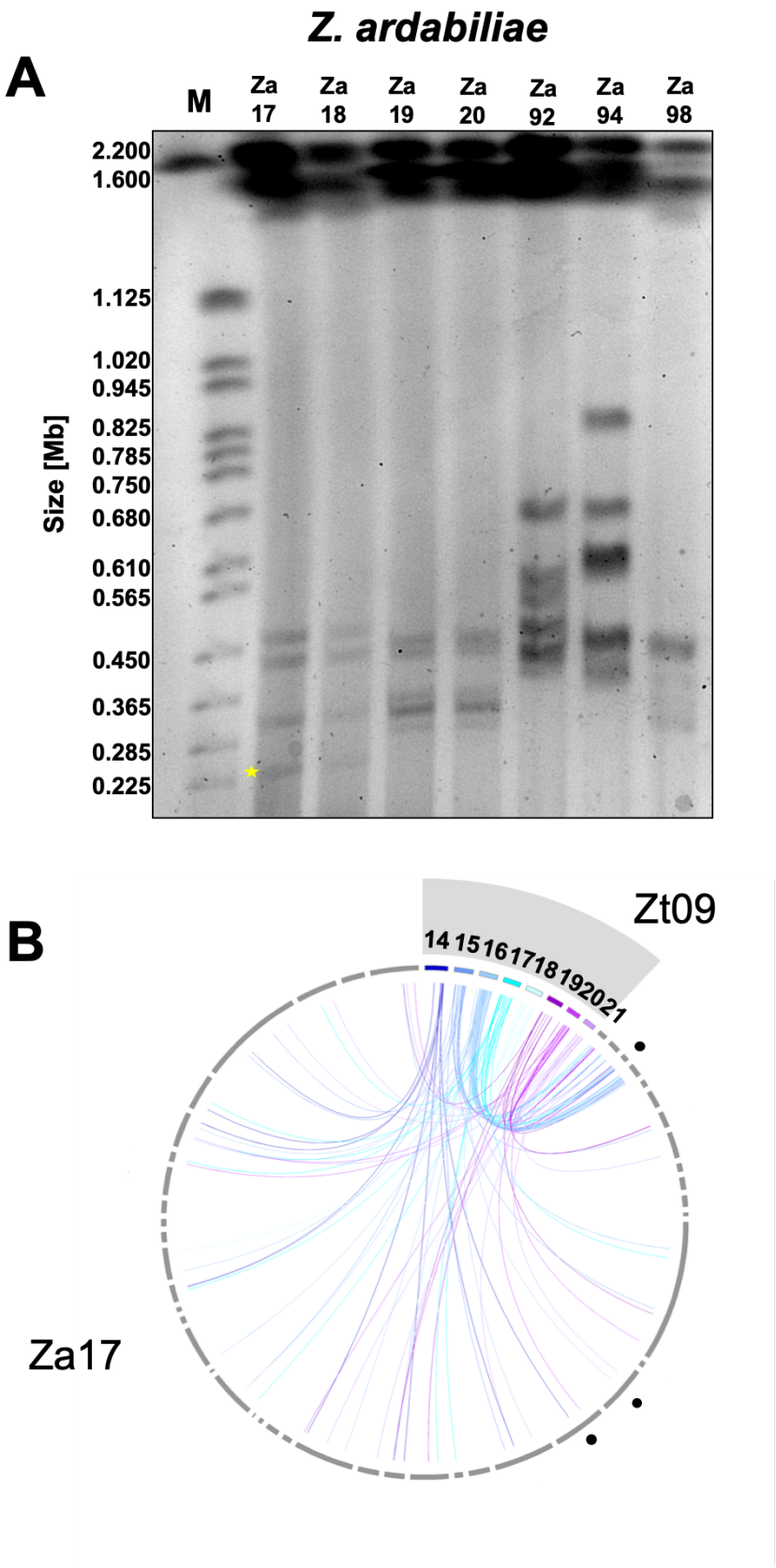


Figure 27 - Presence of accessory chromosomes in *Z. ardabiliae*. A) The pulsed-field gel electrophoresis showed length and size polymorphisms of the smallest fraction of chromosomes with lengths shorter than 2.2Mb. The yellow star marks chromosomes that have a fully sequenced length equivalent in the *de novo* genome assembly. B) Comparative analysis of the synteny between the accessory chromosomes of *Z. tritici* with the genome sequence of *Z. ardabiliae*. The results indicated the presence of the accessory sequences in *Z. ardabiliae*.

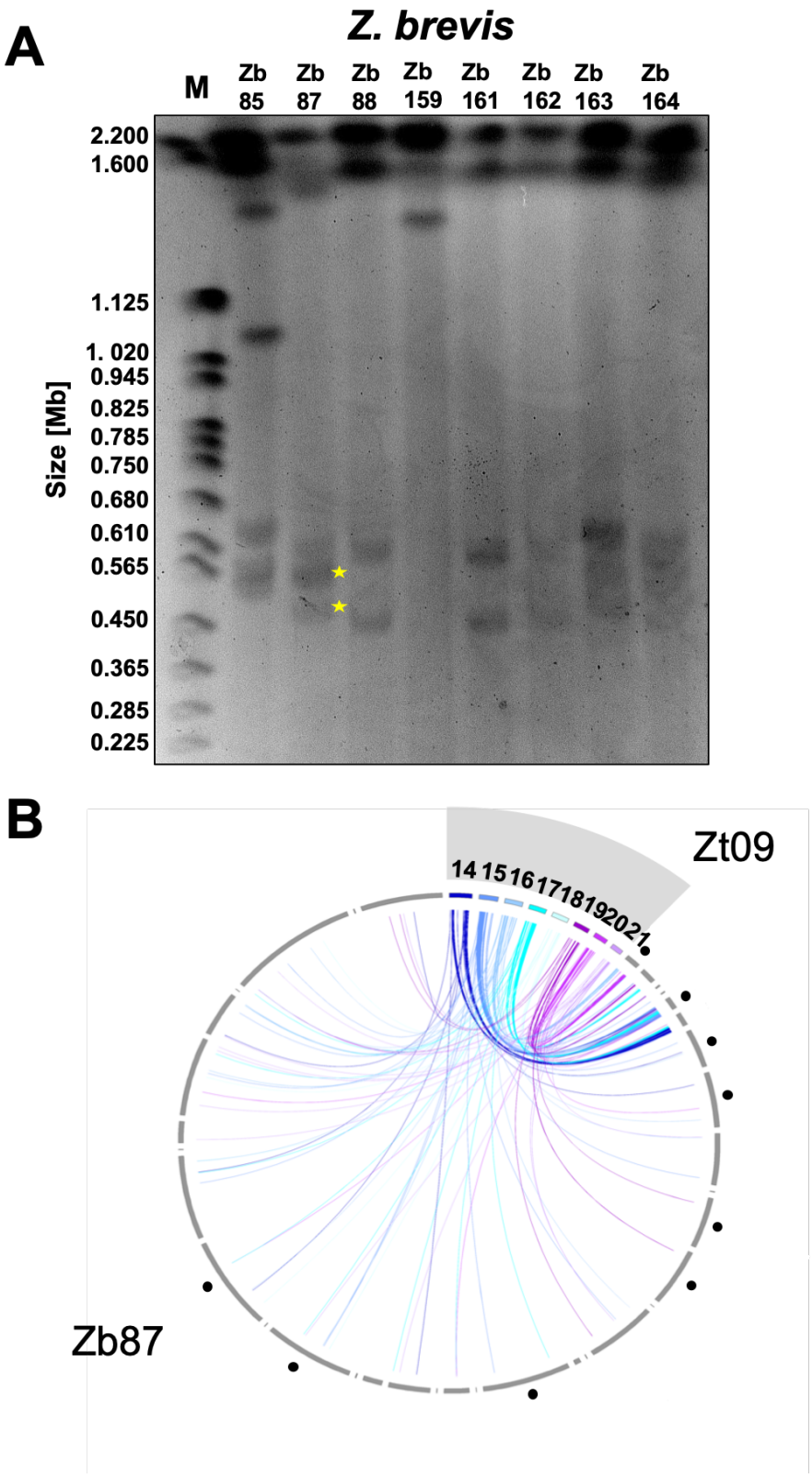


Figure 28 - Presence of accessory chromosomes in *Z. brevis*. A) The result of the pulsed-field gel electrophoresis revealed length and size polymorphisms of the smallest fraction of chromosomes with lengths shorter than 2.2Mb. The yellow star marks chromosomes that have a fully sequenced length equivalent in the *de novo* genome assembly. M = Marker (*S. cerevisiae*; BioRad, Munich, Germany). B) Comparative analysis of the synteny between the accessory chromosomes of *Z. tritici* with the genome sequence of *Z. brevis*. The results indicated the presence of the accessory sequences in *Z. brevis*.

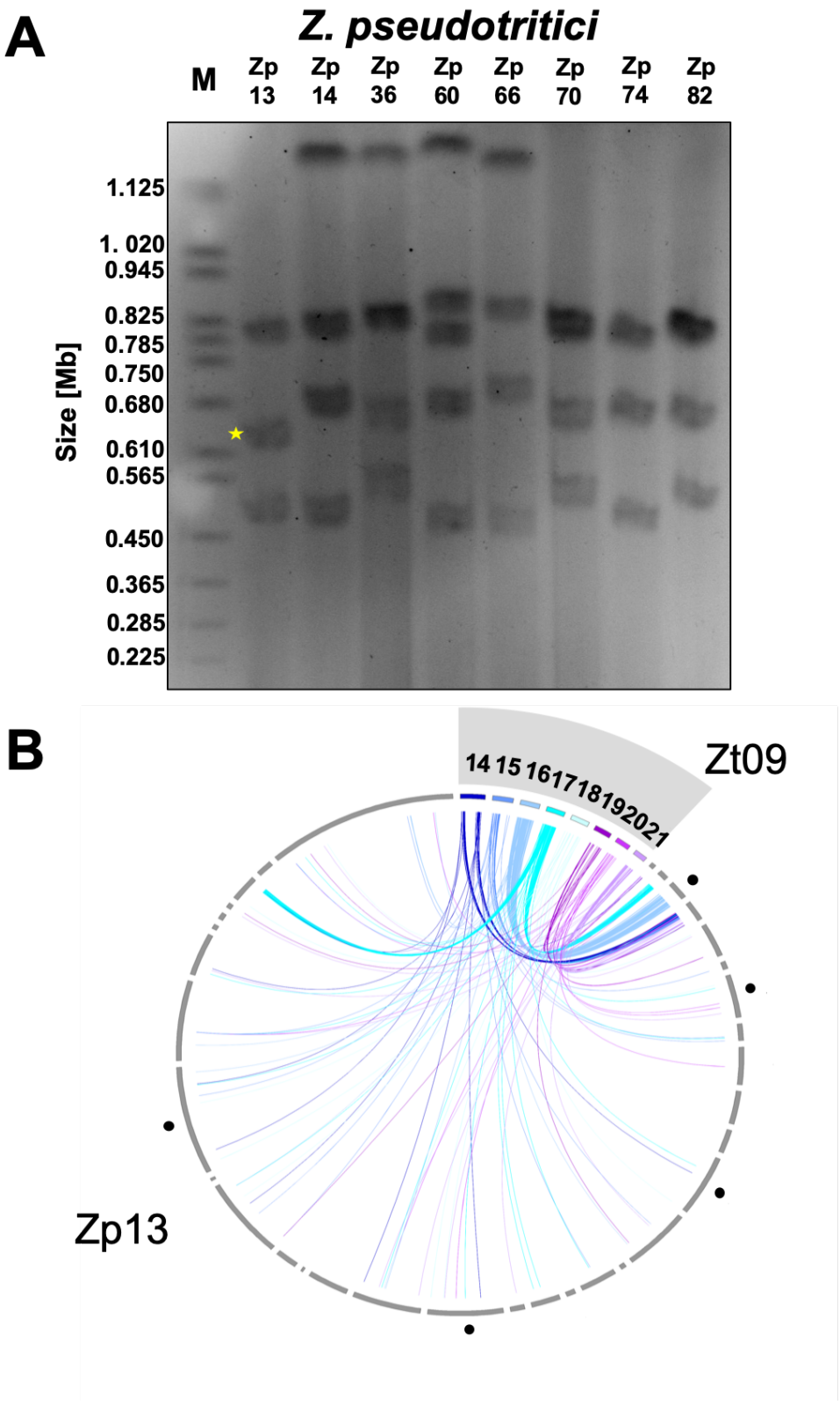


Figure 29 - Presence of accessory chromosomes in *Z. pseudotritici*. A) Pulsed-field gel electrophoresis revealed length and size polymorphisms of the smallest fraction of chromosomes with lengths shorter than 2.2Mb. The yellow star marks chromosomes that have a fully sequenced length equivalent in the *de novo* genome assembly. M = Marker (*S. cerevisiae*; BioRad, Munich, Germany). B) Comparative analysis of the synteny between the accessory chromosomes of *Z. tritici* with the genome sequence of *Z. pseudotritici*. The results indicated the presence of the accessory sequences in *Z. pseudotritici*.

4.2.2. The *Zymoseptoria* species complex showed macrosynteny and a correlation of synteny breaks and transposable elements

Assembly quality and phylogenetic relation affected the pairwise alignments

General statistics of the pairwise genome alignments provided an overview of the general sequence conservation in the *Zymoseptoria* species (see Materials and Methods 3.3.3.1). The results showed a rapid decrease in the proportion of aligned sequences depending on the distance of query and reference in the phylogenetic tree (Figure 25, Figure 30). However, the proportion of sequences aligned against the genome reference of *Z. passerinii* (Zpa63) was much lower (14.8-23.4%) compared to the remaining alignments (44.5-94%) (Figure 30). As a result, the data sets used to perform synteny analyzes showed a tremendous difference in size between the species.

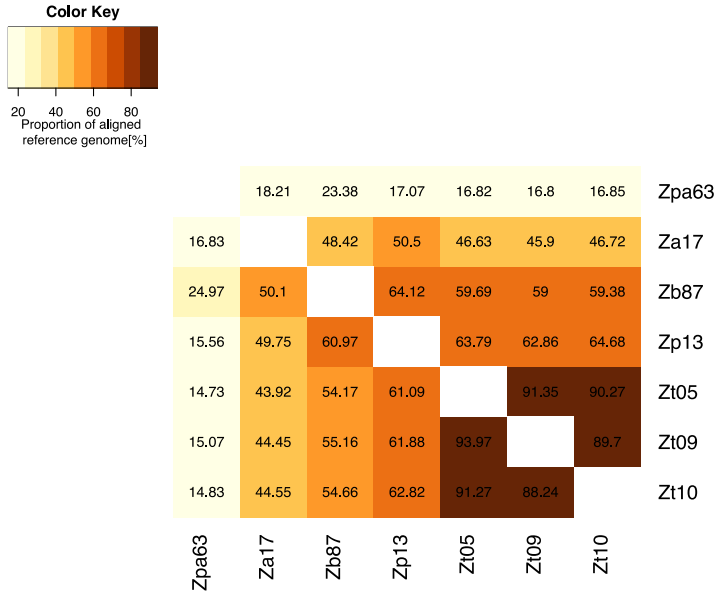


Figure 30 – Proportion [%] of aligned sequences per pairwise genome alignment. The genome of the barley pathogen *Z. passerinii* showed highly reduced coverage with aligned sequences of the remaining sister species than any other pairwise combinations. It is striking that the proportion increased with closer proximity in the phylogenetic tree.

Synteny in the Zymoseptoria species complex

Circular and linear representations of pairwise genome alignments revealed a difference in synteny in the *Zymoseptoria* species complex. General alignment statistics are included in the supplement (Figure S 20).

The circular diagrams and dot plots of the comparative synteny analysis (see Materials and Methods 3.3.3.2) showed that the strength of synteny increased with closer positioning in the phylogenetic tree. So not only, as mentioned earlier, the total length of

the alignments increased, but the circle and dot plots showed higher synteny between more closely related species (Figure 31-33 A, Figure S 8 - 13). However, the comparative analyses detected large inter-specific structural rearrangements on three contigs (unitig_0, unitig_1, unitig_18, and unitig_106) in *Z. ardabiliae* (Figure 31 A). The observed rearrangements on contigs 1, 18, and 106 were homologous to sequences of Zt09 chromosomes 1, 4, and 8.

Interestingly, the observed rearrangements located at the ends of the sequences. In the case of unitig_8, the rearrangement indeed contained a higher proportion of telomeric repeats, which might be due to either a past chromosomal break followed by repair or it may reflect a technical bias such as miss-assembly. However, the other large rearrangements were not related to the positions of telomeric repeats. Whether miss-assemblies caused the observed differences or represented real synteny breaks could only be verified by molecular methods such as, for example, gel extraction followed by sequencing or Southern blotting (Southern 1975). Interestingly, a small number of core chromosomes of Zt09 had fully assembled homologous contigs in *Z. ardabiliae* (Figure 31 A), *Z. brevis* (Figure 32 A) and *Z. pseudotritici* (Figure S 6). Besides the large structural rearrangements between *Z. ardabiliae* and *Z. tritici* all inter- and intra-specific comparisons showed small structural rearrangements (Figure 31-33 A). Due to the purging of the set of repetitive sequences used for synteny analysis, eliminated transposable elements as a direct cause of the observed rearrangements.

The synteny dot plots supported this result, which showed small structural rearrangements such as translocations and inversions on the inter-species level (Figure 31-33 B, Figure S 5 - 7). Interestingly, the intra-specific comparisons between isolates of *Z. tritici* also showed a large number of small structural rearrangements (Figure 33 B). The comparison between Zt09 and Zt10, for example, detected a large number of smaller translocations and inversions (Figure 33 B). However, the synteny analysis indicated macrosynteny in the genus due to the central diagonal (Figure 31-33 B). The centered diagonal became more consistent with increasing assembly quality (fewer contigs and higher N50). In general, the results showed that the *Zymoseptoria* species complex, although generally having strong synteny, still has a considerable amount of smaller rearrangements between and in the species. The following section further quantified the synteny by the detection of synteny blocks and breakpoints and associated with the repeating sequence sections.

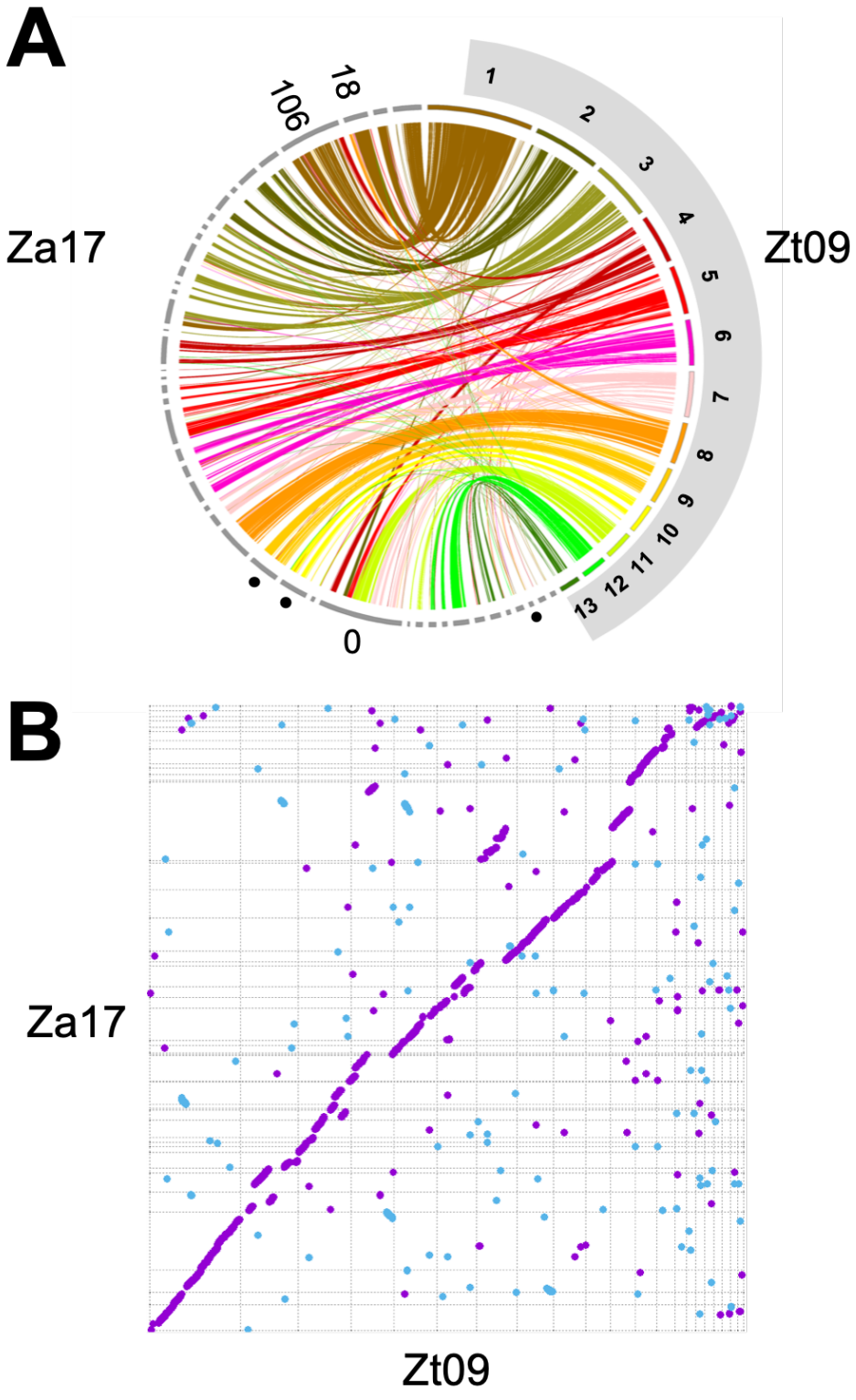


Figure 31 – Synteny plots between *Z. ardabiliae* and *Z. tritici*. A) The pairwise genome comparison revealed a high degree of synteny with indications of sequence conservation of two full chromosomes (*Z. tritici* chromosomes 7 & 8). The results showed three larger rearrangements and many small rearrangements. Black points mark a contig as completely assembled (telomeric repeats at both ends). B) Synteny dot plot shows a main diagonal in the middle indicating macrosynteny. Everything off the diagonal marks a translocation. Small rearrangements and translocations were shown to be both inversions and translocations. Different colors indicate location on different strands (blue = inversed; purple = regular).

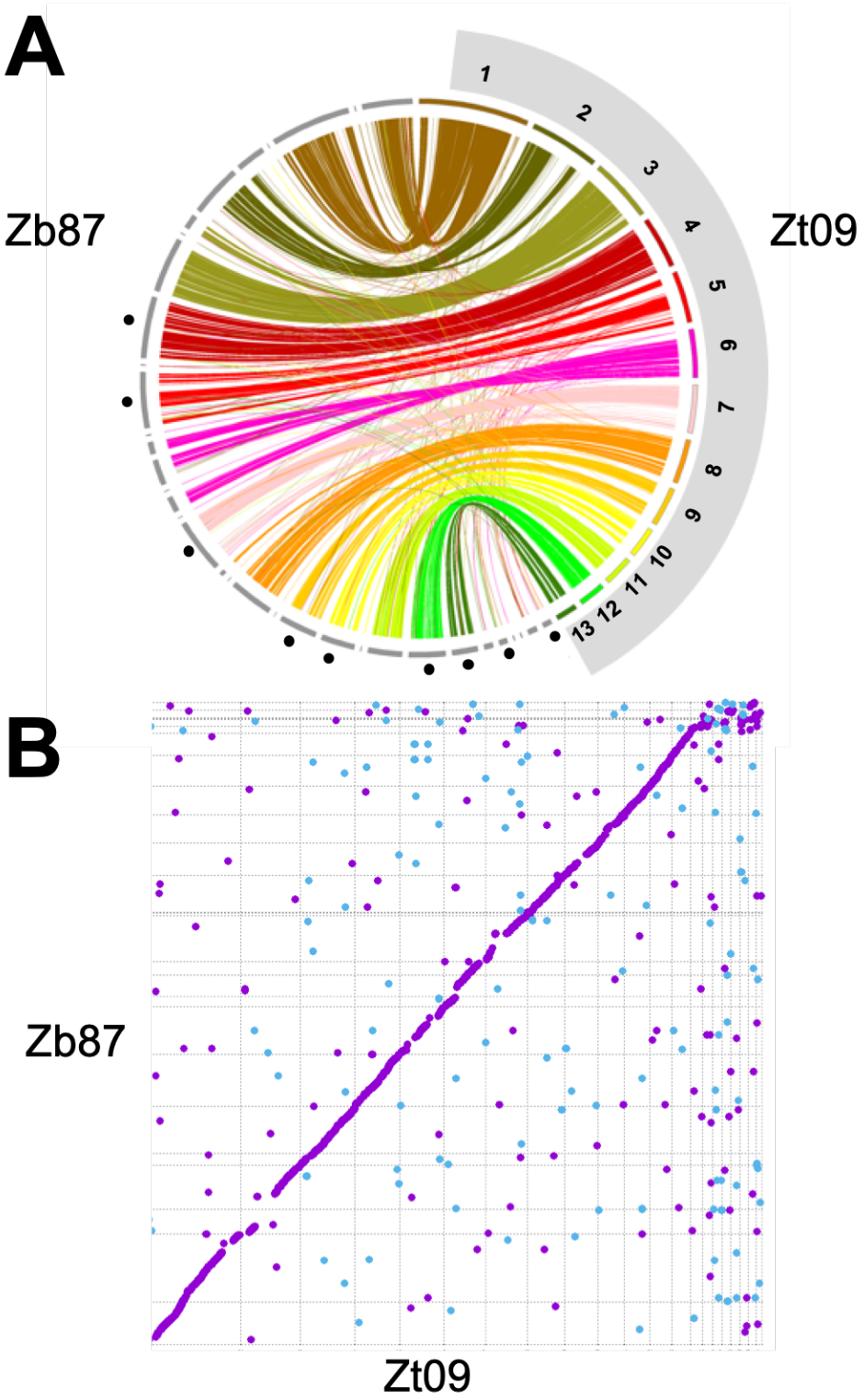


Figure 32 - Synteny plots between *Z. brevis* and *Z. tritici*. A) Pairwise genome comparison revealed a high degree of synteny between Zb87 and Zt09. Interestingly, seven chore chromosomes of Zt09 show completely sequenced homologous contigs in Zb87. Black points mark a contig as completely assembled (telomeric repeats at both ends). B) Synteny dot plot shows a main diagonal in the middle indicating macrosynteny. Everything off the diagonal marks a translocation. Small rearrangements and translocations were shown to be both inversions and translocations. Different colors indicate location on different strands (blue = inversed; purple = regular).

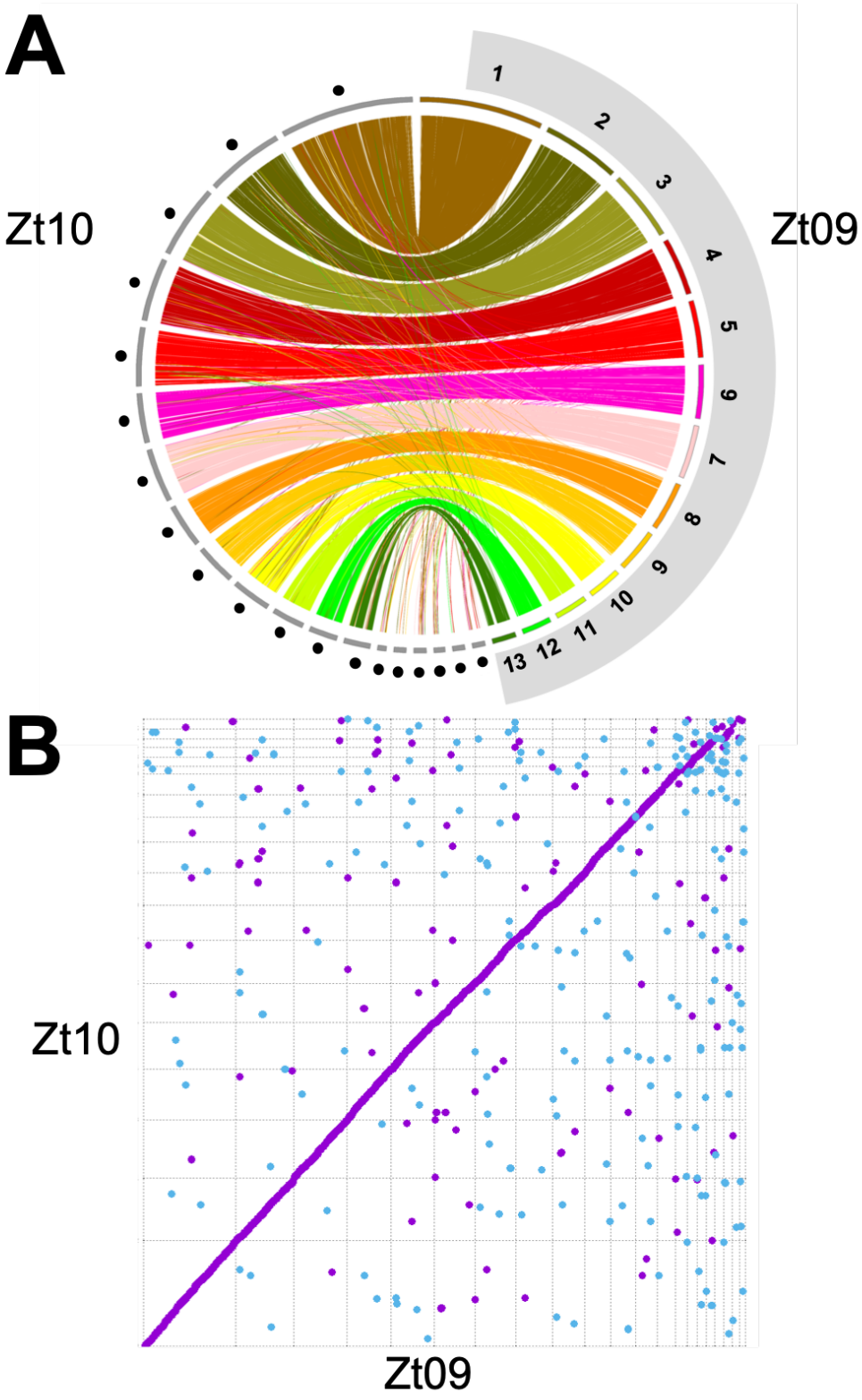


Figure 33 - Synteny plots between two *Z. tritici* isolates. A) Pairwise genome comparison revealed a high degree of synteny between Zt09 and Zt10. Black points mark a contig as completely assembled (telomeric repeats at both ends). B) Synteny dot plot shows a main diagonal in the middle indicating macrosynteny. Everything off the diagonal marks a translocation. Small rearrangements and translocations were shown to be both inversions and translocations. Different colors indicate location on different strands (blue = inversed; purple = regular).

4.2.3. Synteny blocks significantly correlate with the position of repetitive sequences

The number and length of synteny blocks depend on the phylogenetic relationship

Based on the observation of many small translocations on both the interspecies and the intraspecies level, the detection of the total number of synteny blocks quantified the occurrence of structural rearrangements (see Materials and Methods 3.3.3.3). An example of the distribution of the detected synteny blocks in a section of a contig of Zb87 (unitig_62) showed more and shorter synteny blocks in Zpa63 compared to the other species, a trend also observed on the genome-wide scale (Figure 34).

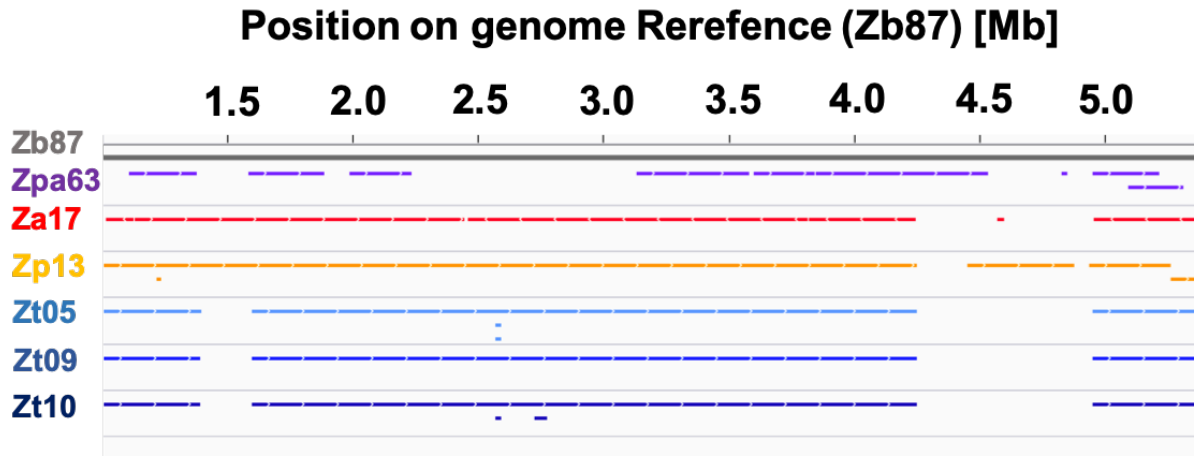


Figure 34 - Example for synteny blocks detected between pairwise comparisons of *Zymoseptoria* genomes. The figure exemplarily represents all synteny blocks between unitig_62 of the reference genome Zb87 (*Z. brevis*) and the query sequences.

Genome-wide, the number, and length of synteny blocks changed with increasing distance in the phylogenetic tree in favor of closely related species/isolates (Figure 35). As measurements for the quantification served the mean block length and the number of synteny breaks per one million base pairs. Thus, intra-specific comparisons in *Z. tritici* showed the most extended and least synteny breaks, while comparisons between distantly related species show an opposing tendency (Figure 35 B). A more significant number of contigs could lead to a higher number of synteny blocks and thus also a higher number of detected synteny breaks, which could have led to the observed high levels (Figure 35 A). However, since the *Z. passerinii* (Zpa63) assembly contained much more contigs, it should be noted that the greatly differing results of comparisons with this pathogen may be explained by this large number of contigs. How far the different results for Zpa63 depended on assembly quality will have to be determined by optimization of the assembly strategy in the future.

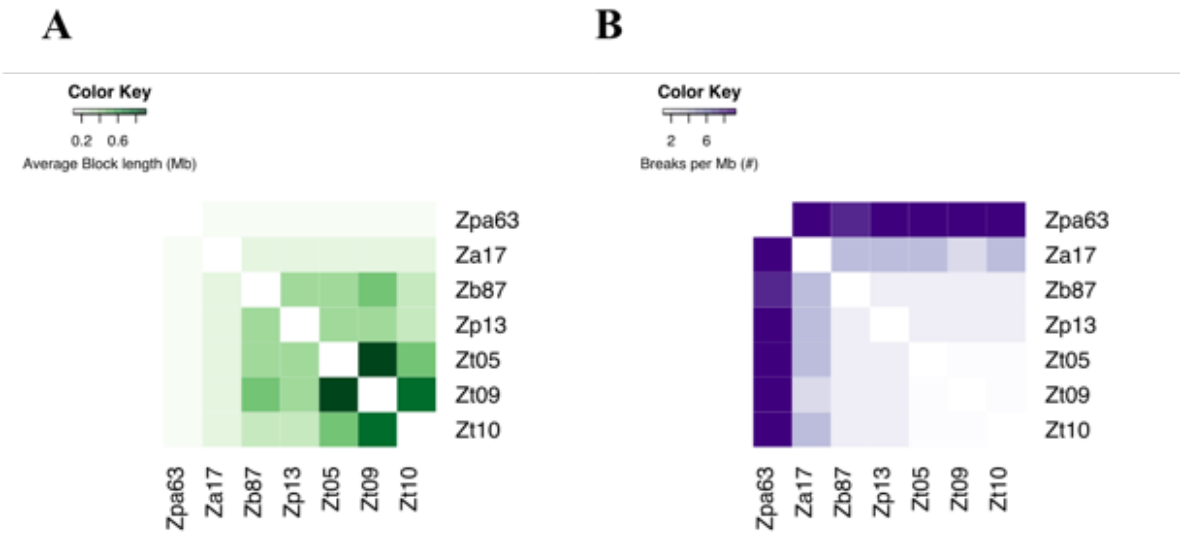


Figure 35 – General synteny block statistics. A) The average block length of the detected synteny blocks. B) The number of synteny blocks per one million base pairs (Mb).

Synteny breaks correlated with the position of repetitive elements

In order to determine whether the positions of synteny breaks was related to the differences in the repetitive element content, the positions of the synteny breaks and known repetitive elements were compared and examined for possible correlation (see Materials and Methods 3.3.3.4). The pairwise synteny breakpoints were compared to known positions of repetitive sequences in the reference genome to test for significant correlation using a permutation test. The majority of comparisons indicated significant correlations between positions of repetitive elements and the pairwise synteny breakpoints (Figure 36). The correlation of repetitive sequences and the pairwise synteny breakpoints between Zt09 and Zpa63 did not show significant correlation (p-value = 0.18) at a significance level of 0.05. The comparisons between Zt09 and Za17 (p-value = 0.003), Zb87 (p-value = 0.003), Zp13 (p-value = 0.001), Zt05 (p-value = 0.001) and Zt10 (p-value = 0.001) showed significant correlation between the positions of synteny breaks and repetitive sequences (Figure 36). However, the correlation analysis was also done with each of the other *Zymoseptoria* reference genomes (Table S 3). The results of these comparisons showed the same pattern of no correlation in pairwise comparisons including Zpa63 but the positive correlation in all other comparisons. The only exception was the comparison between the repetitive genome content of Zp13 and the synteny blocks between Zp13 and Zpa63, which showed a positive correlation between the positions of repetitive elements and synteny breaks. However, these results indicated a correlation between synteny breakpoints and the position of repetitive elements.

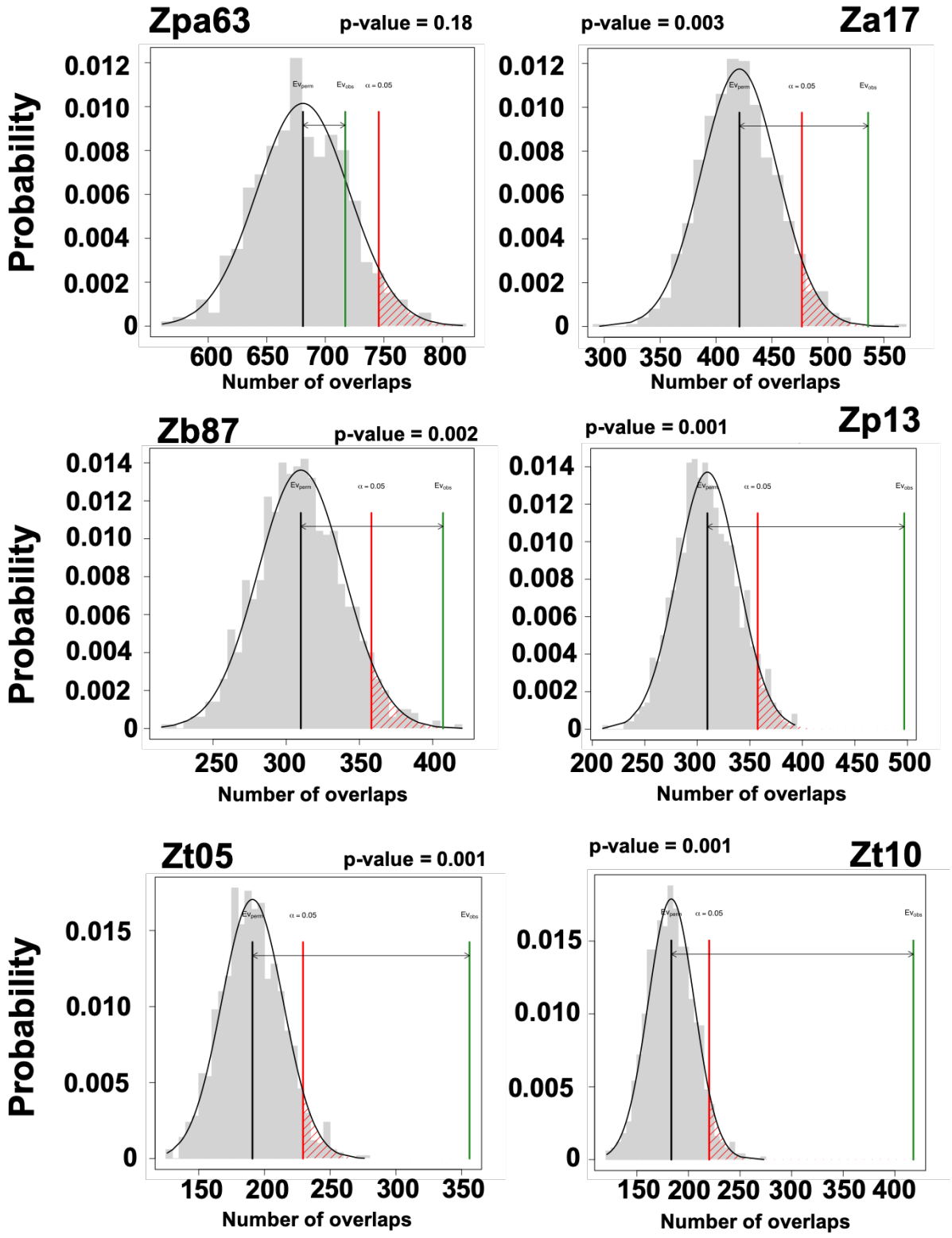


Figure 36 - Pairwise synteny breaks between a reference genome of Zt09 and the other *Zymoseptoria* genomes tested for correlation with positions of repetitive elements using permutation testing. The red line indicates the significance level. The green line represents the observation. An observation beyond the significance level indicates a significant correlation between synteny breakpoints and the positions of repetitive elements. The upper left or right corner of each graph shows the name of the reference strain. The results of the correlation analyses between all other possible pairwise combinations are included on the attached USB-stick (8.15).

4.3. Factors impacting genetic diversity

4.3.1. Patterns of genetic diversity indicate differences in population structure and demography in the *Zymoseptoria* species complex

Structural variation was found to be higher in a pathogen specialized to a managed ecosystem

As the results of the pairwise comparative genomic analyses indicated, smaller structural rearrangements were also present at the intra-specific level of *Z. tritici*. To analyze whether the intra-specific structural variation is equally present in all *Zymoseptoria* species a short-read alignment approach based on discordantly mapped sequencing reads was used (see Materials and Methods 3.3.4). For this, an alignment approach that combined several alignment software aligned intra-specific short-read sequencing data of 16 *Z. ardabiliae*, 8 *Z. brevis*, 24 *Z. pseudotritici*, and 12 *Z. tritici* isolates with *de novo* assembled genomes of Za17, Zb87 and Zp13 and the published genome of Zt09 (Goodwin et al. 2011). The results indicated that the wild grass associated fungal plant pathogens *Z. ardabiliae*, *Z. brevis*, and *Z. pseudotritici* showed less structural variation compared to the wheat pathogen *Z. tritici* (Figure 37). When comparing the differences in proportions of intra- (rearrangements within a chromosome i.e. inversions, insertions, deletions or translocations) and inter-chromosomal (rearrangements between chromosomes i.e. translocations) variation the results showed that the genomes of *Z. brevis* (30.1%) and *Z. pseudotritici* (21.9%) contained a smaller proportion of intra-chromosomal rearrangements compared to *Z. tritici* (42.5%) and *Z. ardabiliae* (44.6%). The isolates of *Z. pseudotritici* tested in this study showed a surprisingly small number of structural variants (352) compared to the sister species. However, the wild grass pathogens, in general, showed very different amounts of structural variants. The pathogen *Z. brevis* (1,210) for example, had only about half of the amount of structural variation that was found in the closely related sister species *Z. ardabiliae* (2,465).

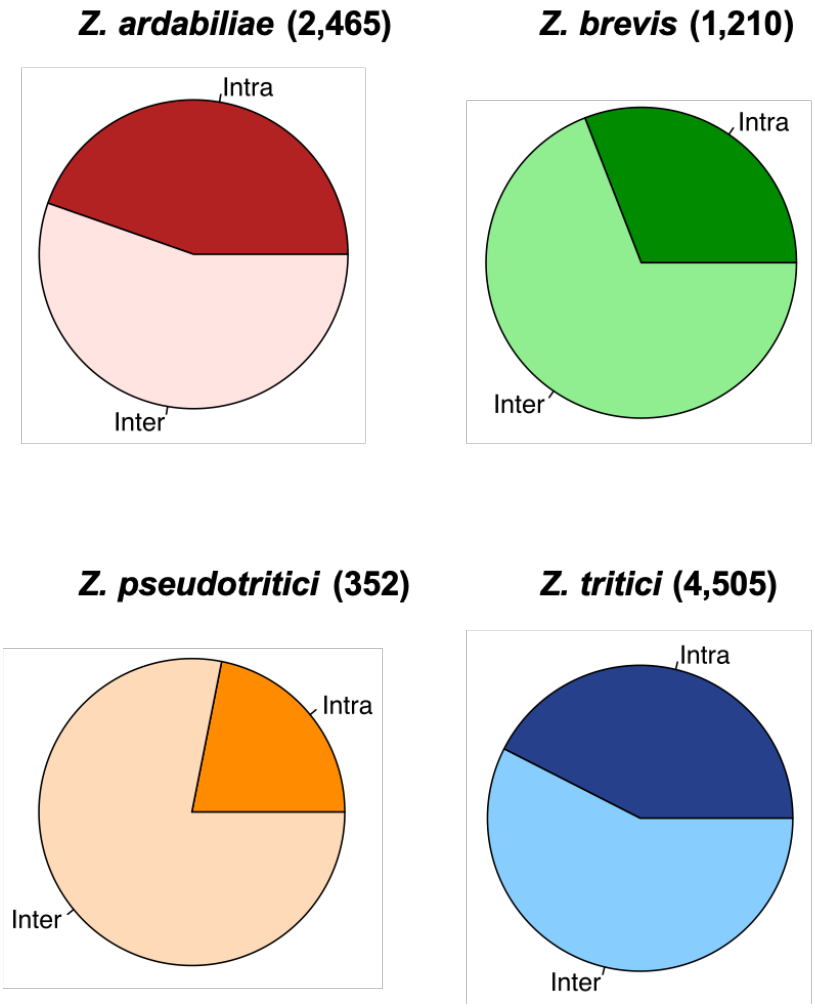


Figure 37 - Structural Variation called by Hydra-Multi. The numbers in brackets after the species name are the total number of structural variants that were detected.

The results showed that the amount of structural variation differs between the *Zymoseptoria* species. However, the differences in the number of genome-wide structural variations indicated different evolutionary histories of phytopathogenic fungi. In order to gain further insight into the possible causes of the differences observed, further genetic diversity studies were carried out using genome-wide SNPs. For this purpose, intra-specific MGAs of *Z. ardabiliae* (N=17), *Z. brevis* (N=9) and *Z. tritici* (N=13) were created and used for the detection of SNPs (see Materials and Methods 3.3.6). Reference projection with subsequent gap filtering reduced the total alignment lengths from 51.6-94.27Mb to 26.47 – 27.27Mb. As a consequence of removing all alignment blocks with less than the maximum number of genomes of the alignment, the variant calling procedure used only the core genomes of the three species. General statistics of the raw and filtered MGA can be found in the supplement (Figure S 21 -22).

The observed numbers of SNPs in the MGAs of *Z. ardabiliae*, *Z. brevis,* and *Z. tritici* showed the number of intra-specific SNPs, and the number of structural variation shows a similar extent of polymorphisms. More precisely, after removing all sites harboring missing data the wild grass associated pathogens *Z. ardabiliae* (979,017) and *Z. brevis* (387,821) showed fewer SNPs, compared to the highly specialized wheat pathogen *Z. tritici* (1,412,138). Interestingly, the number of SNPs detected in *Z. brevis* was three times smaller than the number of SNPs in *Z. ardabiliae* (Table 5). As structural and nucleotide variation both revealed differences between *Z. ardabiliae*, *Z. brevis* and *Z. tritici*, a first survey of the nucleotide diversity and site frequency spectra in these species should give the first hint how demographic events might have shaped genetic variation. Population genomic methods calculated the intra-specific variation using two estimators of nucleotide diversity Tajima's π and Watterson's θ (Tajima 1989), describing the population genetic parameter 4Neμ (Tajima 1983, 1989; Watterson 1975) (see Materials and Methods 3.3.6). As expected, the global average of θ and π was smaller in *Z. brevis* compared to the other grass pathogens *Z. ardabiliae* and *Z. tritici* (Table 5). The wheat pathogen *Z. tritici* showed nucleotide diversities twice as high than the sister species *Z. ardabiliae*, which might be an indication of accelerated evolution or an increase in the effective population size due to the adaptation to an agricultural ecosystem.

Table 5 – Intra-specific sequence diversity statistics. The table reports the total number of SNPs (#), two estimates of the nucleotide diversity (Tajima's Pi & Watterson theta), the average nucleotide diversity calculated as $(\theta_\pi + \theta_W)/2$ and a test of neutrality (Tajima's D).

Species	SNPs [#]	Tajima Pi (π)	Watterson theta (θ)	Nucleotide diversity [∅]	Tajima D
Z. ardabiliae	979,017	0.012	0.0098	0.011	0.49
Z. brevis	387,821	0.005	0.0058	0.005	-0.55
Z. tritici	1,412,138	0.021	0.0197	0.020	-0.07

Therefore, the next step analyzed if changes in demography caused the differences in nucleotide diversity found between the species, so changes population size (see Materials and Methods 3.3.6). As θ and π behave differently under the presence of selection, their difference was used by statistical tests as Tajima's D to test if sequence sets fit the standard neutral model (Tajima 1989). When applied on a genome-wide scale this test statistic reflects the effects of demographic changes (Maruyama and Fuerst 1985). Since it uses the number of segregating sites instead of the number of pairwise

mismatches, the calculation of θ is more affected by changes in the number of low-frequency alleles than the calculation of π. This results in a shift of the test statistic Tajima's D towards positive or negative values range (Tajima 1989). Thus, recent population size expansions may result in a genome-wide negative Tajima's D value due to an excess of low-frequency alleles (Tajima 1989; Slatkin 1996). Bottlenecks lead to the reduction of low-frequency alleles, which then shifts the genome-wide Tajima's D towards positive values (Tajima 1989). The genome-wide inference of Tajima's D showed very different results in the tested *Zymoseptoria* species (Figure 38 A). A genome-wide average close to zero, as found for *Z. tritici*, may indicate a panmictic and neutrally evolving population (Table 5). Both *Z. ardabiliae* and *Z. tritici* showed a uni-modal distribution of values, whereas the distribution of Tajima's D showed a bi-modal shape in *Z. brevis* (Figure 38 B). The genome-wide positive Tajima's D value in *Z. ardabiliae*, potentially resulting from the reduction of the low-frequency alleles, may indicate a recent population contraction. The genome-wide positive value in *Z. brevis*, on the other hand, reflects an excess of low-frequency alleles, potentially due to population expansion.

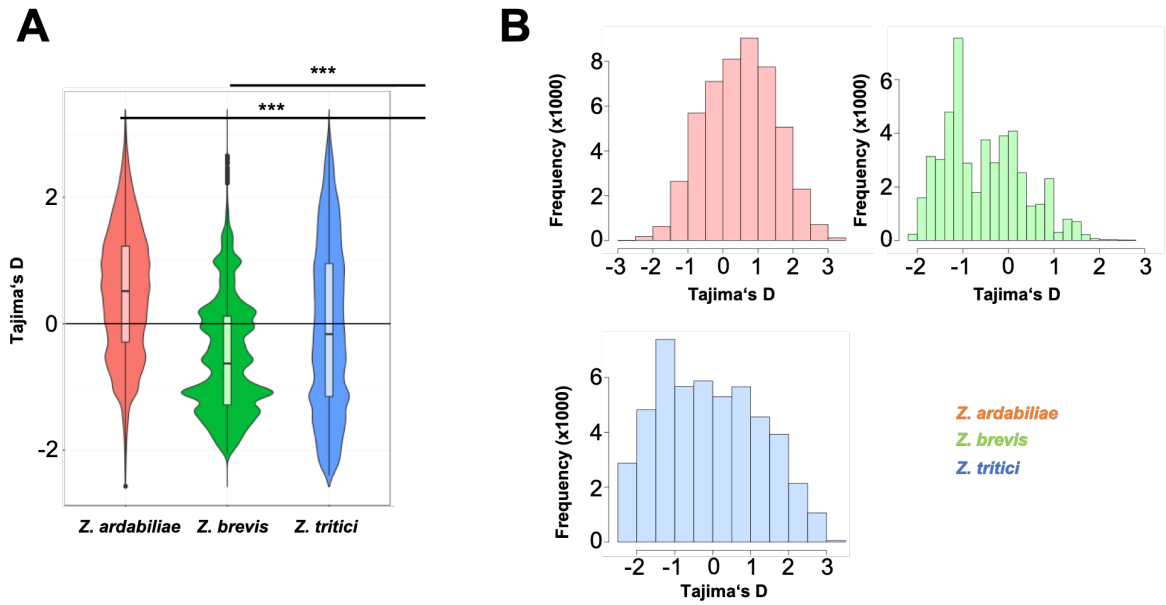


Figure 38 - Genome-wide test of neutral evolution using Tajima's D. (A) Violin plots of the genome-wide inference of Tajima's D using 500bp windows. Statistical significance tested using a Wilcoxon-Wilcox test (*** = 2,2 * 10^{-16}). (B) Histograms show the distribution of the Tajima's D values.

An illustrative and representative result of chromosome 9 in *Z. tritici* and the corresponding homologous regions on untig_35 and unitig_21 in *Z. ardabiliae* and *Z. brevis* showed similar distributions of the estimators of nucleotide diversity θW and π along the genome in all three species (Figure 39). This effect remained constant for all

chromosomes and untigs in all species, as indicated by the data (8.15). Consequently, different demographic histories have likely influenced the evolution of the Zymospetoria species studied here.

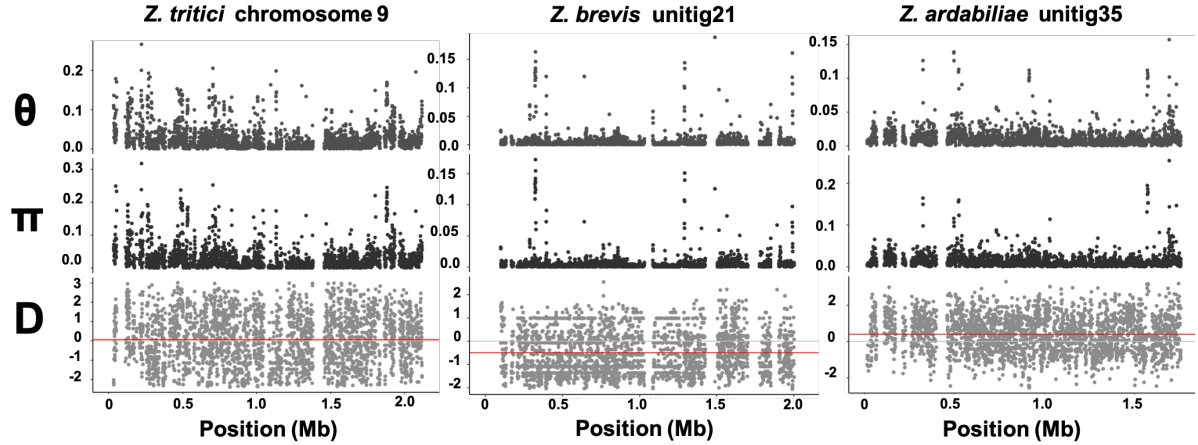


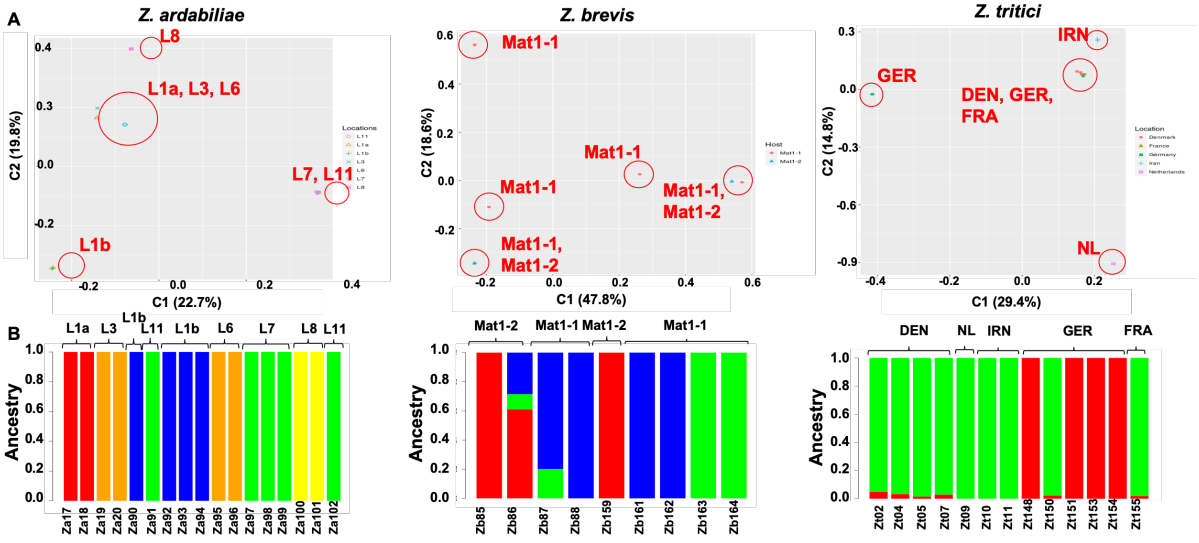
Figure 39 - Detailed illustration of the nucleotide diversity estimates (θ_W and θ_π) and Tajima's D along chromosome 9 of *Z. tritici* and the corresponding homologous sequences on unitig_35 and unitig_21 in *Z. ardabiliae* and *Z. brevis*. The red line indicates a value of zero in the Tajima's D graph. The complete dataset can be found on the attached USB-stick (8.15).

4.3.2. Population structure shapes the genetic diversity in the *Zymoseptoria* species complex

Strong population structure in Z. ardabiliae but weak in Z. brevis and Z. tritici

Pruning the genome-wide SNPs for indels, positions of bad quality and variant pairs in high linkage disequilibrium was conducted before analyses of population structure (see Materials and Methods 3.3.7). This resulted in a final dataset of 901,154 (*Z. ardabiliae*), 166,490 (*Z. brevis*) and 1,009,390 (*Z. tritici*) nucleotide variants. The software Admixture 1.3.0 (Alexander et al. 2009) inferred the population structure in the *Zymoseptoria* species, here treated as single populations. Further, the procedure performed a principal component analysis (PCA) on the same dataset. In the dataset of *Z. ardabiliae*, the results of Admixture favored a population size k of five populations with no signs of admixture, whereas the PCA approach suggested four populations using two components explaining 42.5 % of the genetic diversity in the dataset (Figure 40 A & B). In both analyses, the population clusters fitted the sampling locations (Figure 40). However, the results showed no connection between population clusters and host species (

Figure S 25). In a second closely related wild grass pathogen *Z. brevis* the admixture analyses showed population numbers (k) of three (Admixture) and four (PCA) to be most likely (Figure 40 A & B). Observed population substructure could be neither explained by the mating type locus (Figure 40 A & C) nor the original host species (Figure S 26). However, the analysis indicated population admixture between the subpopulations (Figure 40 B). The dataset of the highly specialized wheat pathogen *Z. tritici* was found to be sub-structured into two populations with admixture by Admixture and into four distinct groups by PCA when applying the two most dominant components that together explain 42.2 % of the genetic variation observed (Figure 40 A & B). Most of the European isolates grouped in two distinct clusters representing one particular cluster with four German isolates, a mixed cluster with isolates from France, Denmark, and Germany (Figure 40 A). The Iranian isolates Zt10 and Zt11 grouped in an exclusive cluster, whereas Zt09 (Netherlands) did not cluster with any of the other isolates. The admixture analysis did not show an effect of sampling locations on the grouping into two populations (Figure 40 B).



C

Location	Year	Isolates
L1a	2004	Za17, 18
L3	2004	Za19, 20
L1b	2011	Za90, 92, 93, 94
L6	2011	Za95, 96
L7	2011	Za97, 98, 99
L8	2011	Za100, 101
L11	2011	Za91, 102

Mating type locus	Isolates
Mat1-1	Zb87, 88, 161, 162, 163, 164
Mat1-2	Zb85, 86, 159

ID	Origin	Isolates
DEN	Denmark	Zt02, 04, 05, 07
FRA	France	Zt155
GER	Germany	Zt148, 150, 151, 153, 154
IRN	Iran	Zt10, 11
NL	Netherlands	Zt09

Figure 40- Population structure analyses in the *Zymoseptoria* species complex. A) Principal component analyses. The two components explain 42.5% (*Z. ardabiliae*), 66.4% (*Z. brevis*) and 42.2% (*Z. tritici*) of the observed genetic diversity in the species. Red circles indicate groups of individuals and red text parts refer to the sampling location (*Z. ardabiliae*, *Z. tritici*) or the mating type locus (*Z. brevis*). B) Admixture analysis of the *Zymoseptoria* species. The entries above the graph indicate the origins *(Z. ardabiliae*. *Z. tritici*), and the original host. C) Detailed information which individual belongs to which location (*Z. ardabiliae*. *Z. tritici*) or host (*Z. brevis*).

Population structure contributed to the observed patterns in Tajima's D in Z. ardabiliae

Population genomic analyses of sub-populations in *Z. ardabiliae*, defined based on the results of the PCA, assessed the effect of population admixture on the distribution of LD and nucleotide polymorphisms (see Materials and Methods 4.3.2, Table 6). The wild grass associated pathogen *Z. ardabiliae* separated into three larger populations that contained individuals from the sampling locations L7 and L11 (Pop1), L1a, L3 and L6 (Pop2) and L1b (Pop3). The analyses of the population structure in *Z. ardabiliae* indicated no admixture between the subpopulations (Figure 40 B). To test if the genome-wide positive average in Tajima's D (Figure 38 A) may be the result of population structure another calculation of genome-wide Tajima's D addressed each of the sub-populations (pooling of Pop1, Pop2, and Pop3) separately.

Table 6 - Sub-Populations of *Z. ardabiliae* used for demographic analyses.

Species	Population	Isolates in population
Z. ardabiliae	Pop1	Za91, 97, 98, 99, 102
	Pop2	Za17, 18, 19, 20, 95, 96
	Pop3	Za90, 92, 93, 94

The genome-wide inference of Tajima's D in the subpopulations of *Z. ardabiliae* showed opposing signatures between the tested populations (Figure 41). A genome-wide negative mean of Tajima's D in the populations Pop1 & Pop3 was detected, while Pop2 showed a positive genome-wide mean.

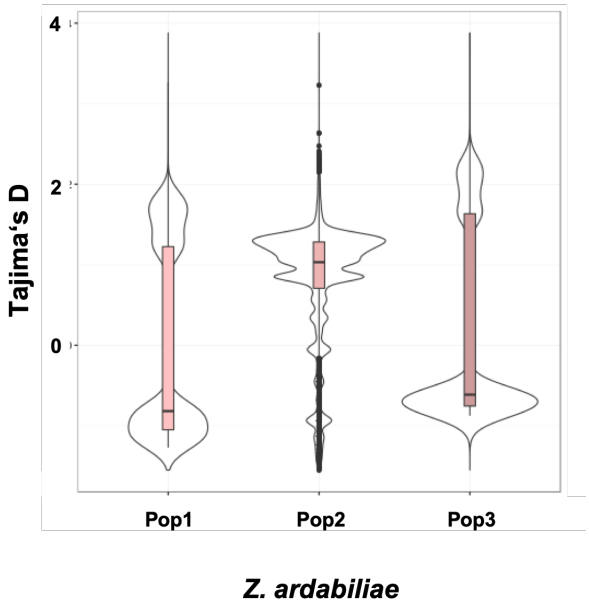


Figure 41 - Genome-wide Tajima's D of sub-populations in *Z. ardabiliae* indicates the bias of the analysis by artificial population admixture. The three sub-populations show opposing signatures of Tajima's D indicating that artificial population admixture biased the overall result.

Demographic effects possibly shaped patterns of genetic diversity in the Zymoseptoria

Because the global inference of Tajima's D indicated the influence of demographic events and population structure a set of analyses investigated their impact on the evolution of the *Zymoseptoria* species complex. As demographic processes could affect the linkage between markers (Hartl and Clark 2007), this Ph.D. study assessed the strength of linkage disequilibrium (LD) in the *Zymoseptoria* species by calculating the correlation coefficient r^2 (see Materials and Methods 3.3.7). The analysis was performed chromosome-wise for each species separately using sliding windows of 50kb on reduced datasets of approximately 38,000 – 42,000 genome-wide SNPs. Because the procedure used for SNP reduction included a random sampling of SNPs, the analysis repeatedly ran 1,000 times per species to ensure that the procedure produced consistent results. One example with an average LD close to the middle of the distribution of mean LDs of the 1,000 (Figure S 27 - 29) underlined a larger LD in *Z. brevis* ($r^2 = 0.62$) compared to *Z. ardabiliae* ($r^2 = 0.22$) and *Z. tritici* ($r^2 = 0.20$) (Figure 42 A). In general, the r^2 passed a lower threshold of 0.2 in distances of 11,531 bp in *Z. ardabiliae*, and 10,053 bp in *Z. tritici* between to SNPs.

However, the screening for orthologs of the mating type loci *Mat1-1*(AF440399) and *Mat1-2* (AF440398) of *Z. tritici* (Waalwijk et al. 2002) in the genomes of the wild grass pathogens indicated the possibility of sexual reproduction in *Z. ardabiliae* and *Z. brevis*. The results showed orthologs of both mating type loci in *Z. ardabiliae* (7x *Mat1-1* and 10x *Mat1-2*), *Z. brevis* (6x *Mat1-1* and 3x *Mat1-2*) and *Z. tritici* (7x *Mat1-1* and 6x *Mat1-2*).

Another cause of the observed differences in Tajima's D as well as LD could be changed in population size history (Hartl and Clark 2007). Here the program MSMC2 (Malaspinas et al. 2016) was used with standard conditions for the estimation of the past effective population sizes over time in the three *Zymoseptoria* species (see Materials and Methods 3.3.7). The effective population size and generation times were calibrated using an of 3.3e-8 mutations/site/cell cycle as estimated for *Z. tritici* by Stukenbrock and co-workers (Stukenbrock et al. 2011). The estimated evolution of population size history over the last 25,000 generations supported the suggestion that observed differences in patterns of genetic diversity between *Z. ardabiliae*, *Z. brevis*, and *Z. tritici* were due to different demographic histories (Figure 43; Figure S 30 - 35).

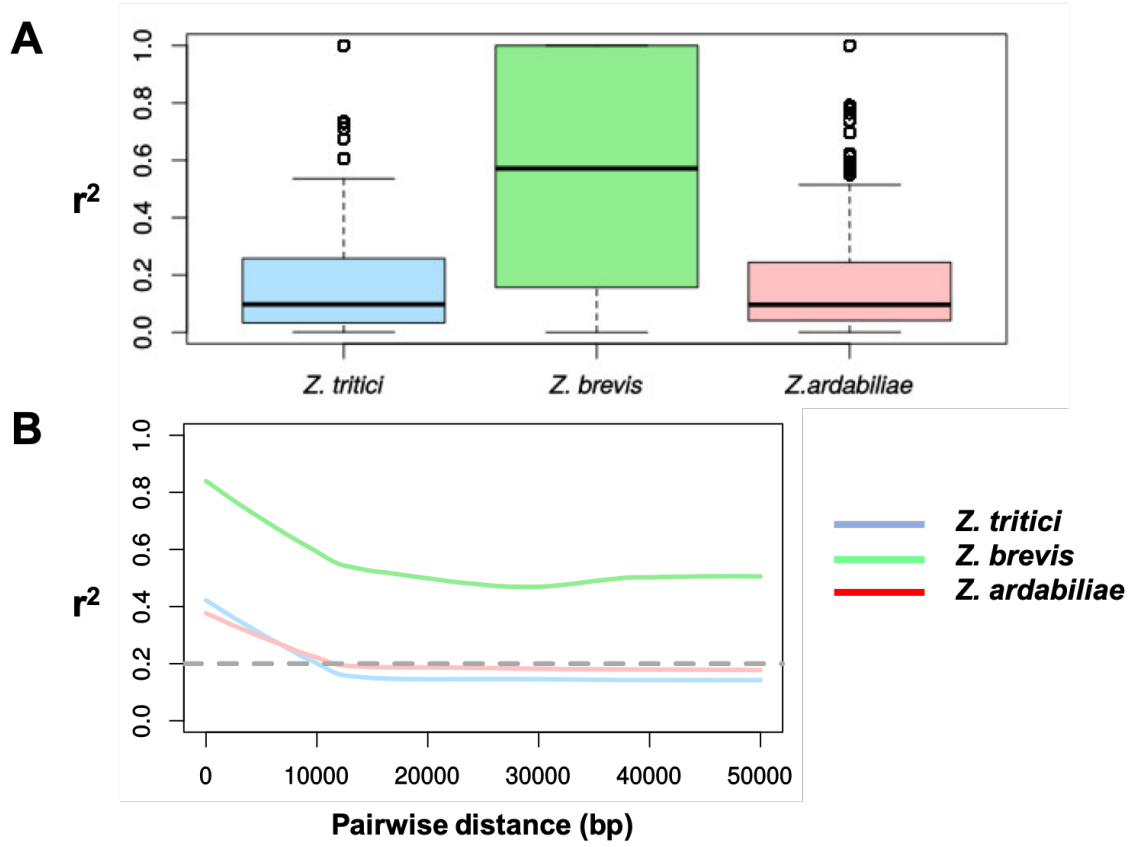


Figure 42 – Linkage disequilibrium in the *Zymoseptoria* species complex. A) Genome-wide average of linkage disequilibrium in *Z. ardabiliae*, *Z. brevis,* and *Z. tritici*. B) The linkage between SNPs and their genomic distance to each other represented using r^2. In comparison to *Z. tritici* (blue) and *Z. ardabiliae* (red), the correlation value r^2 in *Z. brevis* does not fall below 0.2 in a genomic distance of 50 Kb indicating a stronger linkage disequilibrium.

The estimated developments of the population history show a steady increase in the effective population size in *Z. tritici* during the last 5,000 generations. The wild-grass-associated pathogen *Z. brevis* also showed an increase in the effective population size, which started around 2,500 generations ago. The closely related species *Z. ardabiliae* showed a sharp decline in the effective population size (bottleneck) only a few generations before the present. The estimated demographic histories supported a population size expansion in *Z. brevis,* and a recent bottleneck in *Z. ardabiliae* which fitted the results of the genome-wide analysis of Tajima's' D. The estimation of the demographic histories using a different combination of isolates confirmed the results presented here (Figure S 30 - 35).

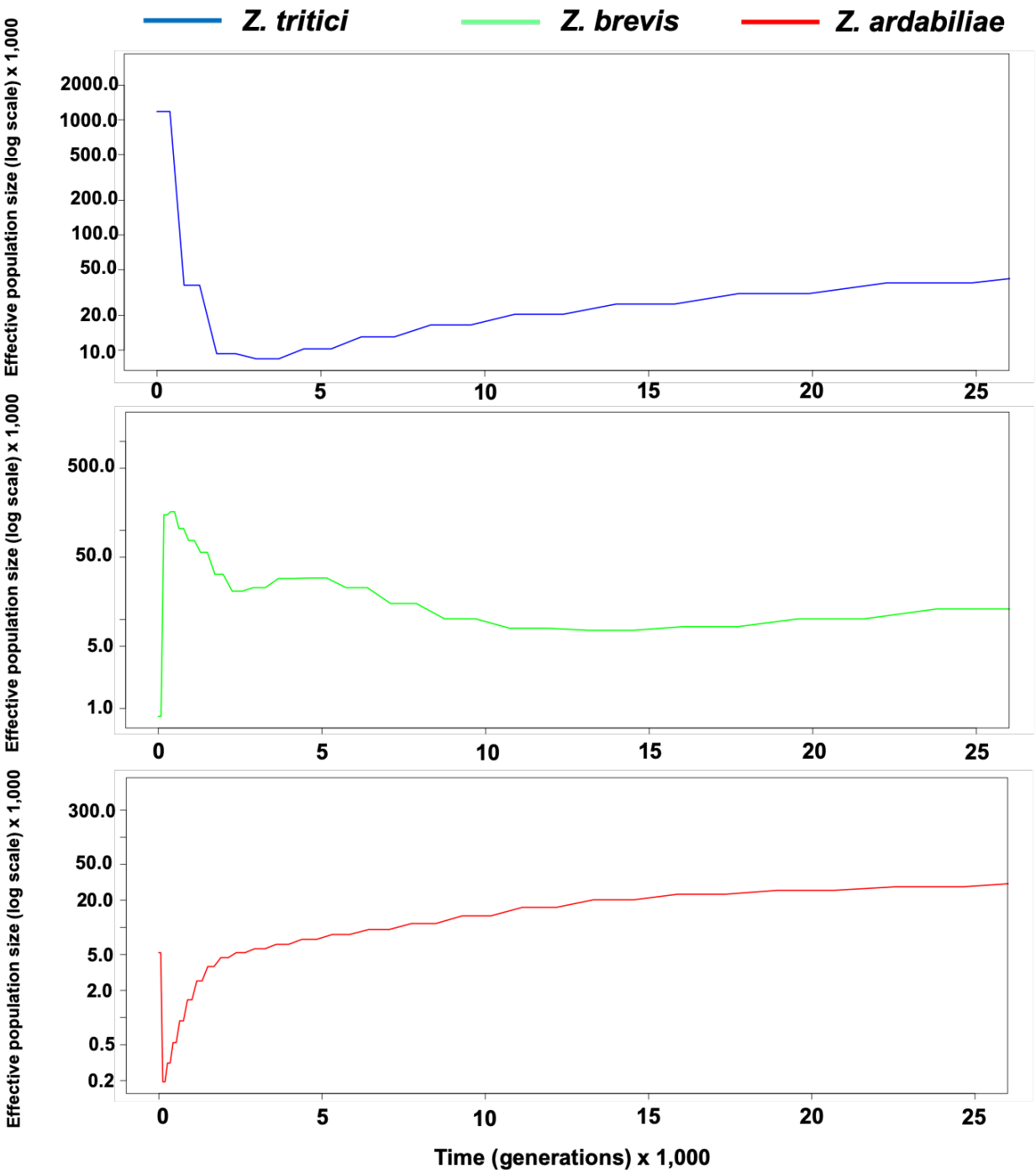


Figure 43 - Effective population size history estimated using the program MSMC2. *Z. ardabiliae* (red) show a steady decrease in contrast to *Z. tritici* (blue) and *Z. brevis* (green) which show population size expansions over the past 2,500 - 5,000 generations. Isolates used for the estimation of the population size history: *Z. ardabiliae* (Za90, 92, 93), *Z. brevis* (Zb86, 159, 162, 164) and *Z. tritici* (Zt02, 04, 05, 10).

4.4. Natural selection and adaptation to different environments

Intra-specific MGAs were screened for genes that are potentially evolving under arms-race dynamics or trench-warfare by testing for traces of past positive directional selection, that might represent speciation and host specialization, as well as signatures of ongoing positive diversifying selection, that might be involved in the contemporary adaptation to the different environmental selective pressures. An inter-specific (McDonald-Kreitman test) and an intra-specific (maximum-likelihood analyses implemented in Codeml) analysis were used (see Materials and Methods 3.3.8). The final number of genes used for subsequent McDonald-Kreitman test (MK test) was less compared to the intra-specific dataset used in PAML (Table 7). A possible reason for this could have been the use of the distantly related outgroup isolate Zpa63 (*Z. passerinii*). Additionally, the strict filtering conditions used, which include gap ratios and sequence number thresholds led to a reduction of successfully tested genes compared to the gene annotation.

Table 7 - Number of genes tested in the MK test and PAML (Yang 2007) analyses. The output files of Codeml for each tested gene can be found on the attached USB-stick (8.15).

		PAML	MK Test
Species	**No. genes**	**No. genes tested**	**No. genes tested**
Z. ardabiliae	10,261*	8,688	3,899
Z. brevis	10,371*	9,007	3,538
Z. tritici	11,839**	9,388	3,935

* = Gene annotation by Sanchez-Ramirez and co-workers (Sanchez-Ramirez unpublished). ** = Gene annotation by Jonathan Grandaubert and co-workers (Grandaubert et al. 2015).

Ongoing intra-specific diversifying selection more frequent in the managed ecosystem

An LRT that compared codon models M1a and M2a, as well as M7 and M8, implemented in Codeml (Yang 1997; Yang 2007) detected genes with traces of intra-specific positive diversifying selection (see Materials and Methods 3.3.8). Diversifying selection is inferred from an excess of non-synonymous over synonymous mutations ($\omega > 1$). With 835 genes in the intercept between the models, more genes showed traces of this type of selection in the wheat pathogen *Z. tritici* compared to the wild grass associated pathogens *Z. ardabiliae* (388) and *Z. brevis* (39) (Figure 44 A). Interestingly, fewer genes evolved under positive diversifying selection in *Z. brevis* compared to its sister species. This result supported the previous signs of increased genetic diversity in *Z. tritici* compared to the closely related sister species. Thus, adaptation to an agricultural ecosystem, such as a wheat field, appeared to lead to the positive diversifying

selection of a large number of genes. However, it is possible that the results in the wild grass pathogens were influenced by the predicted bottlenecks, which led to a reduction in genetic diversity and thus to a lower number of genes with signatures of positive diversifying selection.

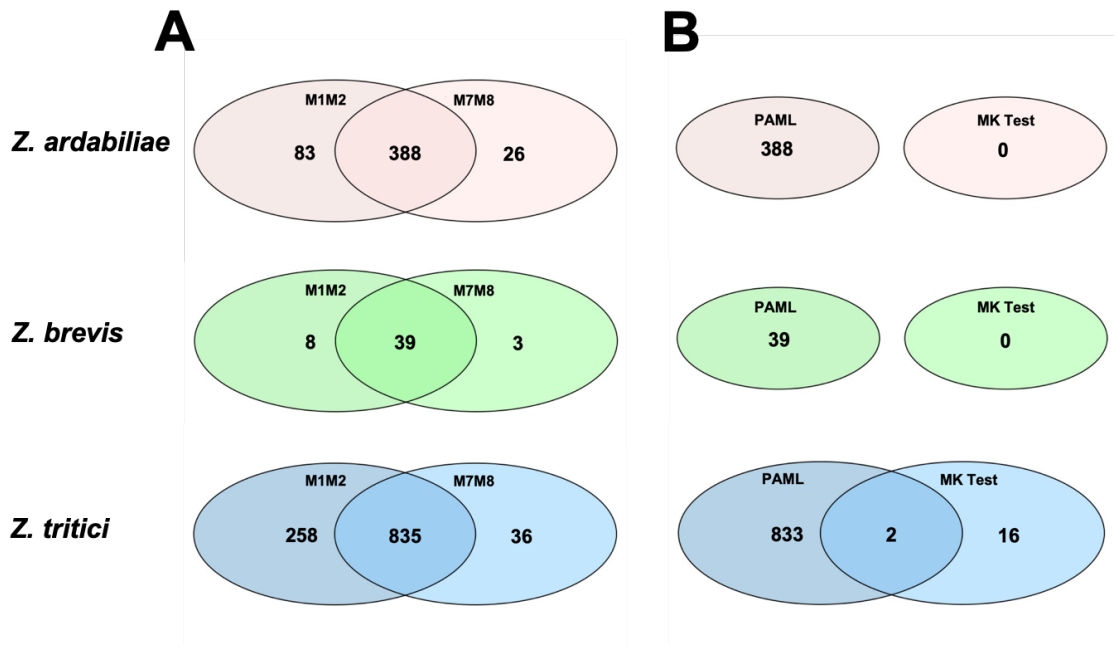


Figure 44 -Number of genes that show signatures of positive and positive diversifying selection. A) Comparison of genes that were estimated to underlie positive diversifying selection. The final dataset was formed by those genes that lay in the intersect of the results obtained from the model comparisons (M1/M2 and M7/M8) tested with PAML (Yang 2007). (8.15). B) Intersect of genes that are driven by positive directional selection (McDonald-Kreitman test) and positive diversifying selection (PAML). Lists of the genes in the intercept as well as lists for the compared datasets are included on the attached USB-stick (8.15).

Inter-specific directional selection only detectable in Z. tritici

Screening for genes that showed signs of positive selection using the MK test showed a similar result (Figure 44 B). In total 18 genes showed signatures of positive selection in the wheat pathogen *Z. tritici* whereas no genes with this signature were identified in the wild grass associated pathogens *Z. ardabiliae* and *Z. brevis*. Having found only signs of positive directional selection in *Z. tritici* could make these 18 candidates attractive candidates of further investigation to understand the genetic adaptation to the domesticated host species, wheat.

Putative ABC-transporter shows signatures of both recent diversifying and past directional selection

The comparative analysis of genes with signatures of positive selection showed that two genes *Z. tritici* showed both signatures of positive diversifying and directional selection (Table 8). One of these genes, *chr_7_000125*, encoded a putative ABC transporter. This possibility made *chr_7_000125* an exciting candidate for further laboratory experiments. The second gene, *chr_10_00386*, encoded a potential component of the fructose metabolism. It could therefore also be an exciting candidate for closer examination of environmental adaptation by *Z. tritici*.

Table 8 - Genes with signatures of positive diversifying (PAML) and directional selection (MK test). aa = Amino acid

Gene ID	Functional prediction	Protein sequence Length [aa]	Mean similarity [%]	GO-terms associated
chr_7_00125	Putative ABC transporter	1,513	88.1	GO:0005524 GO:0005886 GO:0006855 GO:0008559 GO:0016021 GO:0042908
chr_10_00386	Related to 6-phosphofructo-2-kinase 1	876	80.6	GO:0003873 GO:0005524 GO:0006000 GO:0006003 GO:0046835

Since ABC-transporters have a potential role in pathogenicity and fungicidal resistance (Stergiopoulos et al. 2003; Gutiérrez-Alonso et al. 2017) the *chr_7_00125* was further examined to determine if the detected sites under positive diversifying selection belong to the active sites of the protein (Figure 45). A prediction of the functional protein domains showed that the gene codes for a protein (chr_7_00125) with two ABC-transmembrane domains as well as two ATPase domains (Figure 45).

The comparison of sites with signatures of positive diversifying selection (Codeml, M2a & M8) and the positions of the protein domains showed that sites under positive diversifying selection in the first ABC-transmembrane domain and both ATPase-domains (Figure 45). However, further curation of the sites under positive diversifying selection by manual verification in the sequence showed that the site of interest in the first ATPase-domain (position 668) had a conserved amino acid residue which indicated a false-positive result. Additionally, none of the verified sites under positive diversifying selection located in the putative binding sites of the protein (Table S 4, Table S 5).

However, if the sites under positive diversifying selection have another effect on the function of the protein by altering, for example, the protein structure remains unknown.

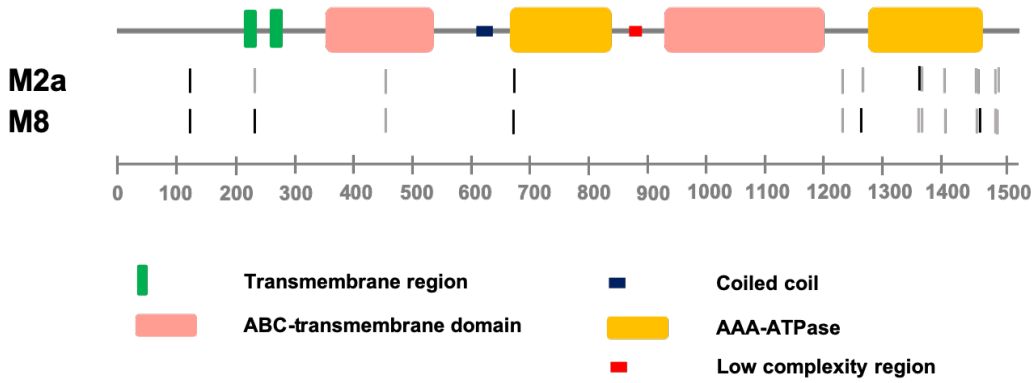


Figure 45 - Predicted protein domains of the putative ABC-transporter *chr_7_00125* as predicted by the SMART web server. The figure additionally shows the positions of the sites found to be under positive diversifying selection as reported by the BEB analysis implemented in the site models M2a and M8 in Codeml. The scale beyond shows the positional information along the protein sequence in amino acids.

Effectors are differently affected by natural selection

For the detection of gene candidates that are potentially involved in virulence, a genome-wide screen searched for effector candidates (see Materials and Methods 3.3.8). In contrast, the analyses of natural selection predicted more effector candidates in the wild grass associated pathogen *Z. ardabiliae* (348 in isolate Za17) than in the closely relates sister species. The analysis found 300 and 172 effector candidates in the genomes of the *Z. tritici* isolate Zt09 (IPO323) and in the *Z. brevis* isolate Zb87, respectively. When comparing the predicted sets of effector candidates with those genes with information about signatures of selection, the results showed, that effector candidates evolving under positive diversifying selection were significantly over-represented in *Z. brevis* while they were significantly under-represented in *Z. ardabiliae* (Table 9). On the other hand, effector candidates of the wheat pathogen *Z. tritici* showed no significant enrichment among the positively diversifying selected genes.

Interestingly, none of the effector candidates, that was also tested in the McDonald-Kreitman analysis revealed any sign of positive directional selection (Table 9). Effector candidates thus appear to have no signatures of past positive directional selection, as would have been expected, for example, in host adaptation in the course of a speciation process (Table 9). However, since the number of effector candidates conserved between the outgroup *Z. passerinii* and the tested *Zymoseptoria* species was meager, the influence of positive selection on species-specific genes was examined in more detail below.

Table 9 - Effectors that evolved under positive diversifying selection and/or adaptive evolution. The total number of effectors is here defined as the total number of effector candidates that were also successfully tested for selection. PDirS = Positive directional selection, PDivS = Positive diversifying selection. The lists of effector candidates as well as of the genes with signatures of positive diversifying selection can be found on the attached USB-stick (8.15).

	Effectors in PAML Dataset		Effectors in MKT Dataset	
Species	**Total**	**PDivS**	**Total**	**PDirS**
Z. ardabiliae	282	6[a]	19	0
Z. brevis	147	4[b]	13	0
Z. tritici	229	18	17	0

a = Significant over-representation of effector-like proteins under positive diversifying selection in *Z. brevis* (p-value = 0.003595; significance level 0.05). Significance was tested using Fisher's Exact test.

b = Significant under-representation of effector-like proteins under positive diversifying selection in *Z. ardabiliae* (p-value = 0.02811; significance level 0.05). Significance was tested using Fisher's Exact test.

Species-specific genes were significantly enriched in effectors and significantly not affected by positive diversifying selection

In total the analysis showed about 2.4x more species-specific genes in *Z. tritici* than in the closely related sister species *Z. ardabiliae* and *Z. brevis* (Table 10). This finding suggested that species-specific genes may play a role in adapting to the agricultural environment. There were significant differences in the frequencies of species-specific genes among the effector candidates and genes under positive diversifying selection in all sister species (Table 10). Effector candidates, for example, showed significant over-representation of the species-specific genes in all *Zymoseptoria* species tested, whereas genes with signatures of diversifying selection were significantly under-represented in *Z. ardabiliae* and *Z. tritici*. Thus, species-specific effectors seemed to play an essential role in adapting to different environmental conditions. Furthermore, diversifying selection seemed to have little effect on the evolution of the species-specific genes.

Table 10 - Over- or under-representation of effector candidates or genes with signatures of positive diversifying selection (PDivS). The resulting gene sets are included as lists on the attached USB-stick (8.15).

Species	Species-specific genes #	Species-specific Effector Candidates	Species-specific Genes under PDivS
Z. ardabiliae	667	135[a]	10[b]
Z. brevis	686	29[a]	5
Z. tritici	1,624	121[a]	64[b]

a = Significant over-representation of species-specific effector like proteins in *Z.ardabiliae* (p-value < 2.2e-16; significance level 0.05), *Z. brevis* (p-value = 2.583e-06; significance level 0.05) and *Z. tritici* (p-value < 2.2e-16; significance level 0.05). Significance was tested using Fisher's Exact test.

b = Significant under-representation of species-specific genes under positive diversifying selection in *Z. ardabiliae* (p-value = 1.065e-05; significance level 0.05) and *Z. tritici* (p-value = 2.103e-08; significance level 0.05). Significance was tested using Fisher's Exact test.

5. Discussion

Since its beginning approximately11,000 years ago, agriculture has become an essential factor in human society (Zeder 2011). Ensuring everyone's access to adequate food security is still one of the biggest agricultural challenges today (Godfray et al. 2010). The effects of climate change and the constant threat of plant pests and diseases are essential fields of current research (Foley et al. 2011; Wollenberg et al. 2016; Dillard 2019). The danger posed by phytopathogenic fungi should not be underestimated, as they are responsible for severe crop failures for many crop species (Fisher et al. 2012). Six out of the ten most dangerous phytopathogenic fungi infect cereals (Dean et al. 2012). Crop cereals have the highest contribution to the global agricultural products which correlates with the high abundance of cereal infecting species among the most dangerous fungal plant pathogens (Dean et al. 2012; FAO 2018)

Pathogens of crops have adapted to environmental conditions that are very distinct from conditions of natural ecosystems. For example, the domestication of wild grasses into cereals and the development of modern agricultural methods such as monocultures, crop rotation, regular fertilization, and the use of fungicides and resistant cultivars have their own Ecosystem (Stoate et al. 2001; Evenson and Gollin 2003; Gurr et al. 2003; Moonen and Bàrberi 2008; Van De Wouw et al. 2010; Pingali 2012; McDonald and Stukenbrock 2016). In order to successfully propagate in these fast-paced ecosystems, it is necessary that pathogens adapt very quickly to new conditions. It is assumed that the speed of adaptation is driven by accelerated evolution (van der Does and Rep 2007). The comparative analysis of closely related pathogens specialized on plants growing in managed, or natural ecosystems provide novel insight into processes of adaptation in agricultural ecosystems of phytopathogenic fungi. Wild grasses often serve as a reservoir for phytopathogenic fungi, from which host-jumps on important cereals can take place. The important rice pathogen *M. oryzae*, for example, arose from a host jump between a wild grass and rice (Maciel et al. 2014).

The work presented here uses members of the *Zymoseptoria* species complex as a model to answer the question of which evolutionary differences have developed in the course of adaptation to agricultural or natural ecosystems. Since the genus *Zymoseptoria* contains both a highly specialized wheat pathogen, as well as several wild grass-associated species, this species complex provides excellent conditions for the comparative analyses between agricultural and natural ecosystems (Stukenbrock, Quaedvlieg, et al. 2012).

One of the main objectives of this work is to investigate whether the genome completeness known from *Z. tritici* (Goodwin et al. 2011) represents a general property of the genus or is species-specific. Furthermore, the research presented here investigates whether the *Zymoseptoria* species have developed differences in their genetic diversity during their evolution and whether these are due to distinct properties in the population structure or the demographic processes. Also, the genomes of selected *Zymoseptoria* species are screened for signatures of positive selection as well as for effector candidates in order to gain a closer look at the host adaptation.

5.1. Genome plasticity is a general trait of the *Zymoseptoria* species complex

Genome plasticity characterizes the genomes of many plant pathogenic ascomycetes, a trait associated with pathogenicity and speciation (Fierro and Martín 1999; Raffaele and Kamoun 2012; Stukenbrock 2013; Galazka and Freitag 2014; Seidl and Thomma 2014). It is known that the characteristics of genome plasticity, such as polymorphisms of chromosome length or number thereof, cause genome length polymorphisms in phytopathogenic fungi (Zolan et al. 1995; Wittenberg et al. 2009; Baxter et al. 2010; Dodds 2010; Duplessis et al. 2011; Croll and McDonald 2012; De Jonge et al. 2013). Insertions, deletions, translocations, the expansion of repetitive elements, as well as breaks or fusion of whole chromosomes may cause length polymorphisms of individual chromosomes (Zolan et al. 1995). The highly specialized pathogen *Z. tritici* exhibits extensive chromosome length polymorphisms at the population level (McDonald and Martinez 1991; Wittenberg et al. 2009).

This research uses comparative genomics methods to study the synteny between reference genomes of seven *Zymoseptoria* isolates from 5 different species. The comparison of the total lengths of the reference genomes used shows similar lengths both within and between the species (39-42Mb). Thus, the values obtained are in the range of the published total length of the reference genome of the *Z. tritici* isolate IPO323 (here referred to as Zt09) (Goodwin et al. 2011). The reference genome is composed of thirteen nuclear chromosomes and eight accessory chromosomes. Accessory chromosomes are characterized by strain-specific presence-absence polymorphisms and have been implicated in genome length polymorphisms (McDonald and Martinez 1991; Dean et al. 2005; Cuomo et al. 2007; Mehrabi et al. 2007). This raises the question of whether the intergeneric genome length polymorphisms between wild-grass-

associated pathogens are also due to accessory chromosomes. Fractions of the genomes of reference isolates of *Z. ardabiliae* and *Z. pseudotritici* have homology with the accessory chromosomes of *Z. tritici* (Stukenbrock et al. 2011). Furthermore, the results of an experimental evolution experiment show that the reference strain Za17 (*Z. ardabiliae*) can lose some of the smaller chromosomes during vegetative growth (mitotic growth), both *in vitro* and *in planta* (Möller et al. 2018). These results suggest the existence of accessory chromosomes in the wild grass pathogens.

In order to investigate presence-absence polymorphisms experimentally, pulsed-field electrophoresis was performed (PFGE). Also, methods of comparative genomics were used to shed more light on the conservation of accessory sequences within the genus *Zymoseptoria*. The experimental analyzes show pronounced presence-absence polymorphisms in the wild grass-associated pathogens. These results suggest that accessory chromosomes are a general property of the *Zymoseptoria* species complex. Furthermore, the comparative analyses of the reference genomes confirm that the accessory chromosomes *Z. tritici* show homologies with the genomes of all tested sister species (Stukenbrock et al. 2011). This finding is another indication that accessory chromosomes are a genetic trait of the genus *Zymoseptoria*.
Furthermore, the results indicate the homology of whole accessory chromosomes between *Z. tritici* and *Z. brevis* and between *Z. tritici* and *Z. pseudotritici*. Interestingly, in the intraspecific comparison between Zt09 and Zt10, there is a hitherto unknown accessory chromosome which has no distinct homologies between the accessory chromosomes described (Goodwin et al. 2011). It can, therefore, be assumed that in the future even more accessory chromosomes will be found due to the increased use of long read sequencing technologies.
This raises the question of why accessory chromosomes are shared over the entire species complex. Accessory chromosomes may have a possible role as an "evolutionary cradle" in which new virulence genes can arise without risking the impairment of essential genes (Croll and McDonald 2012). From this, it can be deduced that the accessory ones offer a fitness advantage. In some phytopathogenic fungi, such as *F. oxysporum*, *N. haematococca*, *L. maculans,* or *Colletotrichum higginsianum*, accessory chromosomes offer advantages such as increased virulence or host specificity (Miao et al. 1991; Coleman et al. 2009; Ma et al. 2010; Plaumann et al. 2018). Thus,

virulence can be given to an avirulent strain of *F. oxysporum* by the transfer of two accessory chromosomes (Ma et al. 2010).

Furthermore, it could be shown that only one chromosome is necessary for the virulence of *C. higginsianum* on *Arabidopsis thaliana* (Plaumann et al. 2018). In contrast, in *Z. tritici* virulence was reported to increase or decrease dependently of the absence or presence of single accessory chromosomes, suggesting that they offer situational advantages (Habig et al. 2017; Fouché et al. 2018; Möller et al. 2018). These advantages and disadvantages may be based on the occurrence of avirulence or virulence genes. Avirulence genes trigger the plant immune defense and thus decrease virulence whereas virulence genes, i.e., effectors, may lead to increased virulence in the host (Petit-Houdenot and Fudal 2017; Zhong et al. 2017). However, no avirulence genes and only a few potential effector genes were described on the accessory chromosomes of *Z. tritici* so far (Goodwin et al. 2011; Grandaubert et al. 2015).

Wild-grass associated pathogens such as *Z. ardabiliae* and *Z. brevis* have to adapt to a broad host range (Quaedvlieg et al. 2011; Stukenbrock, Quaedvlieg, et al. 2012). The most diverse possible distribution of accessory chromosomes within the population could also prove useful, as they may allow, for example, the elimination of individual avirulence factors or virulence factors to infect a variety of host genotypes (Miao et al. 1991; Han et al. 2001; Hatta et al. 2002; Coleman et al. 2009; Ma et al. 2010).

Furthermore, phytopathogenic fungi face a variety of abiotic environmental factors. It has been shown that the wheat pathogen *Z. tritici* and the barley pathogen *Rhynchosporium commune* have a high potential for adapting to temperature differences (Zhan and McDonald 2011; Stefansson et al. 2013). Varying precipitation levels are of particular interest for foliar phytopathogenic fungi such as the *Zymoseptoria* species since increased humidity promotes infection (Shaw 1991). Due to global trade and climate change, plant pathogenic fungi are spreading to new areas more quickly (Bebber et al. 2013). This can be done primarily in host-independent (i.e., water or wind) or host-independent (i.e., trade of host plants) ways (Bebber et al. 2013). It is known, for example, that phytopathogenic fungi follow the temperature optimum (Chen et al. 2011; Lenoir and Svenning 2015). It was also shown that climatic changes could not only lead to migrations but also changes in the hosts (Gange et al. 2011).

To successfully migrate or jump between hosts, it may be beneficial to lose or not possess specific accessory chromosomes. However, such a relationship has not yet been demonstrated and provides scope for further research.

The wheat pathogen *Z. tritici* could also benefit from accessory chromosomes in this way. However, apart from a general relationship between fitness and accessory chromosomes, no further observations have yet been made (Habig et al. 2017; Möller et al. 2018). Therefore, the potential advantages of accessory chromosomes, as well as wild grass-associated pathogens, continue to provide room for speculation. For example, accessory chromosomes may be involved in a variety of different adaptations, such as adaptation to new fungicides or resistant host genotypes.
The results of this thesis support the assumption of the fitness-dependent role of accessory chromosomes. The conservation of accessory chromosomes in the *Zymoseptoria* species complex suggests that under certain circumstances they may offer an evolutionary advantage. Otherwise, it would be likely that due to their cost, they would have disappeared over time due to natural selection. The accessory chromosomes in *Z. tritici* were proposed to be maintained in the population due to a mitotic drive similar to selfish genetic elements known in plants and animals (Jones 2012; Houben 2017; Habig et al. 2018). However, presence-absence polymorphisms may be maintained by frequent losses during vegetative growth (Möller et al. 2018).

The observed genome length polymorphisms cannot be attributed solely to the presence of accessory chromosomes. Chromosomal rearrangements triggered by insertions, deletions, or translocations can also affect the length of a chromosome.
The *Zymoseptoria* species complex was examined for structural rearrangements and synteny, using comparative analyzes. The order Capnodiales, which belongs to the class Dothideomycetes and contains the genus *Zymoseptoria*, is characterized by mesosyteny between species of different genera (Hane et al. 2011; Ohm et al. 2013). Below the genus level, however, the Dothideomycetes show macrosynteny. The results of this dissertation support this finding as it reports macrosynteny within the genus *Zymoseptoria*. However, this result is also reproducible within other classes of Ascomycetes (Shi-Kunne et al. 2018). For example, comparative analyzes between nine *Verticillium* species (Order: Hypocereales; Class: Sordariomycetes) show macrosynteny below the genus level. In contrast, comparisons between two *Aspergillus* species (Order: Eurotiales; Class: Eurotiomycetes; *A. fumigatus* & *A. nidulans*) showed degraded mesosynteny below the genus level (Hane et al. 2011).
The compared genomes of the *Zymoseptoria* species complex showed macrosynteny as well as smaller structural rearrangements. The effects of structural rearrangements

in *Z. tritici* have already been highlighted in several studies (Wittenberg et al. 2009; Croll et al. 2015; Fouché et al. 2018). Cross-breeding experiments of various *Z. tritici* isolates show that chromosomal rearrangements can occur during meiosis (Wittenberg et al. 2009). However, experimental evolution experiments show that large structural rearrangements under stress conditions such as chromosome breaks, fusions or duplications can occur in the genomes of *Z. ardabiliae* and *Z. tritici* (Möller et al. 2018). Also, they frequently occur in the vicinity of transposable elements. The results of this dissertation confirm that this type of structural mutations also occurs in natural populations of the pathogen. Comparative analyses of the reference genomes of the *Zymoseptoria* species revealed a significant correlation between synteny breaks and the positions of known repetitive elements. The relationship between repetitive sequences and chromosomal rearrangements has also been reported in other model systems of phytopathogenic fungi (De Jonge et al. 2013; Faino et al. 2016; Plissonneau et al. 2016). Structural rearrangements can also influence virulence. For example, de Jonge and co-workers reported line-specific regions that are associated with virulence (De Jonge et al. 2013). Structural rearrangements also occur in *Z. tritici*, in this case, close to an effector gene in increased virulence (Hartmann et al. 2017).

Finally, the results presented here show that the *Zymoseptoria* species complex has pronounced genome plasticity with accessory chromosomes and repeat-induced structural variations. Based on previous results from the field of *Zymoseptoria* research as well as other pathosystems, it can be speculated that the observed genome plasticity influences virulence and host adaptation.

5.2. The signatures of intra-specific genetic variation point to differences in population structure and demography

The development of agriculture and the associated domestication of wild grasses resulted in an ecosystem of crops with significantly reduced genetic diversity (Doebley et al. 2006). These ecosystems are characterized by high host densities, low plant diversity and rapidly changing environmental conditions due to anthropogenic impacts (Stahl and Bishop 2000; Linde et al. 2002; Zhan et al. 2002). As a consequence of the increased number of hosts and density, there can be a substantial increase in the effective population size of crop associated pathogens (Stukenbrock et al. 2011). Significant increases or decreases in effective population size may are reflected by the genetic variability (Couch et al. 2005; Ronald and Akey 2005; Zaffarano et al. 2008).

The results of intra-specific population genomic analyses show substantial differences in genome diversity, population structure, demographics between the crop and wild grass-associated pathogens of the genus *Zymoseptoria*. The results further show an increase in both the nucleotide diversity and the structural variation in *Z. tritici*. The highly specialized wheat pathogen has a considerably higher level of genetic diversity than its sister species from natural habitats. One possible cause could be changed in the population size by demographic events. As the population size changes, so do the number of potentially mutating genomes, which also affects the number of low-frequency alleles (Tajima 1989; Slatkin 1996; Ellegren and Galtier 2016). As mentioned earlier, accumulation or reduction of population-related low-frequency alleles have different effects on the calculation of nucleotide diversity estimates π and θ (Tajima 1989; Simonsen et al. 1995; Kreitman 2000; Whitaker et al. 2005; Hartl and Clark 2007). Other parameters, such as changes in the genome-wide distribution of linkage disequilibrium can therefore also be used to infer changes in the effective population size (Hill 1981). The genome-wide patterns of the test statistic Tajima's D (Tajima 1989) in the tested intra-specific datasets of *Z. ardabiliae*, *Z. brevis,* and *Z. tritici* reflect differences in the effective population size. The result of the wheat pathogen thus supports the presence of a panmictic population with random mating and constant population size, while the wild grass associated fungal pathogens indicate significant changes in population size. This finding is consistent with the results of several previous studies, based on RFLP and microsatellite markers that have already highlighted this feature

of the global *Z. tritici* population (Linde et al. 2002; Zhan et al. 2003). The analysis of the global population structure of the wheat pathogen *Z. tritici*, which was carried out in this work, indicated a substructure of the European isolates as well as a single cluster with the Iranian isolates. The substructure described here primarily supports the results of Grandaubert and co-workers, who found a global differentiation into a European and an Iranian subpopulation and a slight substructure in the European isolates (Grandaubert et al. 2017).

The size and the history of the effective population size can be estimated from the distribution of the genome-wide LD. Thus, large effective population sizes may be reflected by a low LD whereas small populations sizes produce a high LD (Sved 1971; Hill 1981; Hayes et al. 2003; Waples 2006; Rogers 2014). Pairwise LD analyzes of SNPs within genomic distances of up to 50 kb support changes in the effective population size in *Z. ardabiliae* and *Z. tritici*. Although the decay of LD in *Z. ardabiliae* is somewhat slower compared to *Z. tritici*, both decays are faster than that of *Z. brevis*. This result indicates a higher frequency of recombination in *Z. ardabiliae* and *Z. tritici* (Nieuwenhuis and James 2016). This result fits the findings of Stukenbrock and Dutheil who found that *Z. ardabiliae* and *Z. tritici* show high recombination rates, whereas that of *Z. tritici* is slightly higher (Stukenbrock and Dutheil 2018). However, the continued decline and flatness of the curve in *Z. brevis* may indicate both a bottleneck in population size as well as a predominantly asexual reproduction (Tenaillon et al. 2008; Rogers 2014). The curve depends strongly on the ratio between random genetic drift and recombination. In cases where the recombination after the bottleneck is stronger than the genetic drift (i.e., in *Z. ardabiliae* and *Z. tritici*), the decay of the linkage disequilibrium becomes steeper.

In contrast, if random genetic drift is stronger than recombination as it might be the case in a clonal population (i.e., *Z. brevis*) the LD decay becomes flat (Rogers 2014). This result contradicts the results of demographic analysis and the global patterns of Tajima D, both of which point to a recent increase in population size in *Z. brevis*. This effect would lead to a much steeper curve than expected (Rogers 2014). The different results could indicate two different scenarios: In the first scenario, *Z. brevis* could predominantly reproduce asexually, which may lead to increased linkage disequilibrium disequilibrium (Prugnolle and De Meeûs 2008). However, according to a second possible scenario, population expansion could be after a recent bottleneck.

The performed Admixture analysis shows that individuals with different mating types share genetic material to varying degrees among themselves. The presence of two mating type loci (*Mat a* and *Mat α*) indicates the possibility of sexual recombination. In *A. fumigatus*, for example, the presence of possible mating types was suggested to indicate sexual reproductive potential (Paoletti et al. 2005). Therefore, sexual reproduction cannot be denied for *Z. brevis*, especially as the analysis of the population structure carried out in this work showed signs of mixing between the isolates. However, the evolution of the population size history, as well as the genome-wide distribution of the parameter Tajima's D, suggest that there has recently been a substantial population increase. However, the population size history does not show any recovery from a population bottleneck. If a bottleneck occurred, it could be due to a drastic reduction in the host population or to a founding effect (Mayr 1954). For example, distinct populations of the plant pathogens *F. graminearum, Rhynchosporium secalis*, and *Mycosphaerella fijiensis* show signs of population expansion after migration-related bottlenecks (Rivas et al., 2004; Zaffarano et al., 2008, Robert et al., 2012, Kelly and Ward, 2018).

In contrast, the population size history of the wild-grass pathogen *Z. ardabiliae* shows a significant bottleneck in the effective population size. Interestingly, in the case of *Z. ardabiliae* both the genome-wide patterns in Tajima's D and the evolution of demographic history point to a bottleneck in population size, while the evaluation of genome-wide LDs does not indicate a bottleneck. However, the opposing Tajima's D values may not be explained solely by demographic changes, as the population structure can lead to genome-wide deviations from neutrality (Hartl and Clark 2007). The observations of this thesis show that the population sample of *Z. ardabiliae* is highly sub-structured. This substructure is more correlated with the sampling site than with the host species. One possible hypothesis that could explain this substructure might be built from the substantial reduction in effective population size. It could be that reduction of the population size might have led to a population sub-structure by reducing the genetic diversity in different local areas differently. By these small local populations with different pathogen genotypes might have been produced that mimic a strong population substructure.

To see if the population structure really could have influenced the results, the three largest subpopulations of *Z. ardabiliae* were tested separately for distribution of allele frequencies. The genome-wide analysis of these subpopulations with Tajima's D

showed a pattern indicating an influence of the population structure on the result. Two of the three subpopulations tested had a genome-wide negative Tajima's D, while the remaining population had a positive mean. Simulations have previously shown that the admixture of populations with contrasting values of Tajima's D leads to an intermediate value (Staedler et al. 2008; Chikhi et al. 2010). Building on this fact, the results of this work suggest that artificial admixture of several subpopulations of *Z. ardabiliae* into a single data set may affect the overall outcome. Simulation studies have shown that population blends could produce comparable signatures as genetic bottlenecks (Flint-Garcia et al. 2003).

Contrary to the genome-wide mean of Tajima's D, the demographic stories of each population point to a recent bottleneck effect. Whether a substantial population expansions in the very near past followed the bottleneck effects, cannot be proven beyond doubt with the methods used here. For example, subpopulations 1 and three could regain their size after the strong bottleneck effect.

However, possible causes of possible population size bottlenecks in *Z. ardabiliae* and *Z. brevis* can only be speculated. However, one possible cause could be the recent severe drought of the Fertile Crescent in the recent past and to this day (Kelley et al. 2015). For around 20 years, moderate to severe droughts have been reported in Syria and Iran (Kelley et al. 2015; Karimi et al. 2018). This persistent lack of water could have led to a collapse of host populations and consequently to a reduction in the population sizes of *Z. ardabiliae* and *Z. brevis*.

In summary, the results not only confirm the panmictic nature of the global *Z. tritici* population but also point to a recent population expansion of *Z. brevis*, the cause of which, however, remains unknown. Furthermore, the *Z. ardabiliae* dataset has a strong substructure, possibly due to a recent strong bottleneck effect.

However, not only selection or demographic events but also hybridization can impact genetic diversity. Population genomic analyses showed a reduction in local genetic diversity along genomes due to hybridization events (Stukenbrock et al. 2012; Menardo et al. 2016). Hybridization, so the interspecific mating due to weak species boundaries is a known source of speciation in fungal plant pathogens (Stukenbrock 2016). The species *Z. pseudotritici* was shown to be a hybrid species (Stukenbrock et al. 2012) Additionally, comparative analyses of the wild grass associated pathogens *Z. ardabiliae* and *Z. brevis* suggested frequent hybridization events (Feurtey unpublished). However, the genome-wide analysis of nucleotide diversity presented here

showed no distinct regions with reduced nucleotide diversity as it could be shown in the case of *Z. pseudotritici* (Stukenbrock et al. 2012). To which extent hybridization had an impact on the genetic diversity reported in this thesis has to be evaluated in future studies.

5.3. Ongoing adaptive evolution and the detection of effector candidates in the genus *Zymoseptoria*

Domestication can directly affect the evolution of phytopathogenic fungi (i.e., *S. cerevisiae*, Duan et al., 2018) or indirect (i.e., *Z. tritici*, Stukenbrock, et al. 2007). Pathogen adaptative evolution in agricultural habitats can be very different from those of natural habitats (Croll and McDonald 2017). Thereby, the use of monocultures can lead to a strong directional selection (Stukenbrock and McDonald 2008). On the other hand, it has also been shown that genes involved in virulence may show positive diversifying selection that increases allelic diversity in the population in response to host and pathogen co-evolution (Aguileta et al. 2009; Stukenbrock and McDonald 2009; Dodds and Rathjen 2010; Sperschneider et al. 2014). Genome-wide screening of fungal pathogens for signatures of positively diversifying selection to find candidate genes that may be involved in host specialization has applied frequently to genome data (Kobmoo et al. 2018; Krishnan et al. 2018). Signatures of a positive selection of genes in the genus *Zymoseptoria* have already been evaluated from comparative analyzes (Stukenbrock et al. 2010; Stukenbrock, Quaedvlieg, et al. 2012). However, the function of the candidate genes was unknown, so gene deletions followed by mutant phenotype assays *in planta* showed the role of the positively selected genes in virulence (Poppe et al. 2015). The main goal of this part of the project was to identify relationships between gene candidates and virulence. In order to determine which genes are particularly subject to accelerated evolution, analyzes included both intra-specific screenings for genes with positive diversification selection signatures and inter-species screenings for positive direction selection genes. This Ph.D. project shows that the wheat pathogen *Z. tritici* has a more significant number of genes with positive diversifying signatures than its close relatives *Z. ardabiliae* and *Z. brevis*. However, Grandaubert and colleagues reported a slightly lower number of genes with diversification-selection signatures compared to the number of candidates found in this thesis (Grandaubert et al. 2017). However, to what extent the detected intrinsic signals of positive diversifying selection are due to the hypothesized different demographic scenarios of wild-grass associated

pathogens is not known at this time. For example, the reduced genetic diversity in the *Z. brevis* dataset could be a cause of the low number of genes with signatures of positively diversifying selection. There are several possible explanations for the excess of genes in diversifying the selection in *Z. tritici*. One possibility would be that the anthropogenically induced rapid changes in environmental conditions (fungicides or resistant varieties) may have led to diversifying selection where a diverse allele pool allows a more flexible adaptation to future challenges (Charlesworth 2006). In addition to effector genes, which will be discussed later in this paper, several studies showed that cell-wall-degrading enzymes (Brunner et al. 2013) and host-specific toxins (McDonald et al. 2013) also evolve under diversifying selection.

Interspecific analyzes using the MK test (McDonald and Kreitman 1991) were used to detect signatures. The result of intraspecific population genomic analyses of targeted selection. Using stringent filter conditions, a small number of genes were predicted to show adaptive evolution in *Z. tritici*, while no genes appeared in the sister species. Since these genes showed signatures of positive selection only in *Z. tritici*, they can play an essential role in adapting to wheat. A comparison of the significantly positive selected genes detected with the PAML and MK tests yielded two genes in both analyzes with a significant result. Another blast search and genome analysis revealed that one of these genes is similar to a putative ATP-binding cassette (ABC) transporter (Accession number: CM001202.1, Goodwin et al. 2011). In fungal pathogens, ABC transporters have previously played a role in virulence as well as in the establishment of resistance to fungicides (De Waard 1997; Schoonbeek et al. 2002; Stergiopoulos et al. 2002) Comparison of sets of genes, however, that developed under both positive selection regimens revealed a potential candidate associated with virulence, a putative ABC transporter.

Interestingly, all of the abovementioned ABC transporters (*MgAtr1* to *MgAtr5* and *MgAtr7*) (Zwiers and De Waard 2000; Schoonbeek et al. 2002; Stergiopoulos et al. 2002) by *Z. tritici* showed no or weak (*MgAtr3*) similarity to the candidate found in this thesis. This finding may indicate that the ABC transporter-like gene found to show signatures of positive selection may be another family of ABC transporters. ABC transporters influence virulence (*MgAtr4*), resistance to fungicides (*MgAtr5*) and resistance to bacterial compounds (*MgAtr2*) (Stergiopoulos et al. 2003; Zwiers et al. 2003) reported new putative ABC transporters might be a good candidate for molecular analysis. He is used *in planta* assays such as gene deletions and mutant phenotyping to

review his role. Interestingly, an interspecific comparative analysis of the genes with signatures of a positive diversifying selection with those without such a signal could not show significant enrichment of species-specific genes. It follows that the majority of genes that may be involved in adapting to the environment are not species-specific. Nevertheless, the species-specific genes, with diversifying selection, may be particularly interesting as candidate genes involved in host adaptation.

Another major class of pathogenicity-dependent genes, termed effectors, often develops through diversification of selection (Martin and Kamoun 2011). These genes encode small secreted proteins and are critical drivers of host specialization (Stergiopoulos and de Wit 2009; Lo Presti et al. 2015). Although fungal effectors have no standard features, general properties such as length, cysteine content, and putative secretion signals are used to predict effector candidates (Rafiqi et al. 2012; Giraldo and Valent 2013; Sperschneider et al. 2016).

The prediction of effector genes in this study shows substantial differences in the numbers of effector candidates between the wild grass associated pathogens *Z. brevis* and its sister species. The reference genome of *Z. brevis* (Zb87) shows only half the number of potential effectors in *Z. ardabiliae* and *Z. tritici*. Comparative analyses of several species of *Colletotrichum* showed that a change in the pathogen-to-endophyte lifestyle was also associated with a decrease in the effector repertoire (Hacquard et al. 2016). Whether this possible explanation also applies to *Z. brevis* is hitherto purely speculative. Interestingly, the comparative analyses of the effector candidates found in this work indicate a conserved accumulation of species-specific effector candidates.

This observation is not unexpected as the loss and gain of effectors play an essential role in host adaptation (Couch et al. 2005). Investigations in the genus *Fusarium* have also led to the conclusion that the fungal species had the vast majority of the effector candidates in common, but the drastic lifestyle differences only a few species-specific effector candidates could have caused these differences (King et al. 2018).

Effector candidates with signatures of diversifying selection may be involved in the evolutionary arms race between host and pathogen (Stukenbrock and Mcdonald 2007; Win and Kamoun 2008; Brunner et al. 2013; Chen et al. 2014). Interestingly, the results of this dissertation show enrichment of species-specific genes among the effector candidates in all three species. These results indicate that species-specific effectors may play an essential role throughout the genus *Zymoseptoria*.

5.4. Limitations of this study

Sample size and the sampling scheme, as well as some method-specific problems, play a significant role in the detection of genetic diversity. The sample size is one of the most critical possible sources of error resulting from the existing data set. In simulations, it was possible to detect an adverse effect of too small sample size on the calculation of the nucleotide diversity and consequently on the calculation of the parameter Tajima's D (Subramanian 2016). For example, a large sample size allows a much more accurate prediction of genetic diversity and population structure (Fumagalli 2013). The results presented here may, therefore, have a significant bias due to the small sample size. In a population genomic study of three closely related species of the fungal mammalian pathogen genus Pneumocystis, similar sample sizes revealed very similar differences in genome-wide nucleotide diversity and Tajima's D in the genus, as in this Ph.D. thesis (Cissé et al. 2018).

Also, it seems possible that the sampling procedure leads to a bias of the results. It, therefore, seems conceivable that the one-time sampling of the *Z. brevis* datasets at one time and a single collection point could have led to a distortion in the form of a reduced diversity compared to the different populations of the other pathogens.

Another potential problem of the work presented here could be the recognition of signatures of a positively diversifying selection with the software PAML. One problem of the models implemented in PAML is that they are based on an underlying phylogenetic tree that does not allow recombination (Anisimova et al. 2001). However, *Z. tritici* and *Z. ardabiliae* have shown a high recombination rate (Stukenbrock and Dutheil 2018). Simulations have shown that although the effects of recombination events are significant, they do not significantly affect the analysis (Grandaubert et al., 2017). For future analyses, similar simulation-based tests should also be performed on all three datasets used in this study to ensure that the *Z. ardabiliae*, as well as the *Z. brevis* record of recombination, are not affected.

6. Conclusion and perspectives

Phytopathogenic fungi like the members of the *Zymoseptoria* species complex are subject to constant selection pressure in the co-evolution with their host and adaptation to their environmental niche. The genus *Zymoseptoria* includes both the highly specialized wheat pathogen *Z. tritici* and several wild grass-associated pathogens, making this species complex an excellent model system for studying host specialization and speciation processes (Quaedvlieg et al. 2011; Dean et al. 2012; Stukenbrock, Quaedvlieg, et al. 2012; Fones and Gurr 2015; Torriani et al. 2015). This study aimed to investigate the differences between pathogens from different ecological environments in terms of genome composition and genetic diversity. This study combined methods of molecular genetics and methods of comparative and population genomics to investigate genome plasticity as well as the effects of population structure and demography using measures of genetic diversity within the *Zymoseptoria* species complex. Furthermore, genome-wide scans of signatures of positive selection and potential effector candidates shed light on the genetic basis of host specification within the species complex. The results of this study confirm a cross-species occurrence of genome plasticity within the genus *Zymoseptoria*. The plasticity is characterized by accessory chromosomes and the appearance of synteny breaks associated with repetitive elements. This observation strengthens the experimental results based hypotheses of a conditional mediated fitness benefit (Habig et al. 2017; Fouché et al. 2018; Möller et al. 2018) conditioned by the accessory chromosomes, as well as the stress-induced instability of the genome (Möller et al. 2018). The results of this Ph.D. thesis suggest that all genomes of the genus *Zymoseptoria* have similar genomic requirements for conditional adaptation to changing environmental conditions.

In this context, the results of the genetic diversity analyze presented here are particularly interesting, as they show that the wheat pathogen *Z. tritici* has a significantly higher effective population size than the closely related wild-grass fungi-associated phytopathogenic fungi. A possible explanation for this observation may be related to evolution within an agricultural ecosystem. As the use of annually recurring monocultures mainly characterizes modern crop cultivation areas, their high host density and the resulting easy transfer between hosts offer excellent conditions for the development and conservation of large population sizes of phytopathogenic fungi (Read 1994). Since the effective population size is directly related to genomic diversity, the rapid increase in this parameter by enabling pathogen adaptations more quickly presents a

big challenge to disease management in agricultural areas. Here, the results of genetic diversity analysis of wild grass associated pathogens, which show substantial reductions and generally smaller or sharply changing effective population sizes. Since natural ecosystems are primarily characterized by lower host densities and higher host diversity, it is natural to consider these attributes when developing new disease management protocols. For example, a move away from monocultures to multi-host pathway strategies with increased host diversity could lead to a reduction in the effective population size and thus the adaptation potential of *Z. tritici*. This measure could reduce both a wide range of airborne pathogens (ascospores) and direct contact between leaves (pycnidiospores) as it limits the number of susceptible hosts available. However, the already high genetic diversity of the pathogen could complicate such countermeasures, since there is already a very high potential for adaptation. This is supported in particular the analysis of positive diversifying selection by the results that show an increased influence of this type of selection in the wheat pathogen *Z. tritici*.
Interestingly, however, this influence cannot be transferred to the effector candidates, i.e., genes that are directly related to virulence. The results of this study show that not positive diversifying selection but rather the occurrence of species-specific effectors may play an essential role in host adaptation. Future experimental analyzes such as gene deletions followed by *in planta* assays are needed to investigate the importance of species-specific effector candidates on virulence further.
However, the detection of signatures of positive diversifying selection not only reflects the increased genetic diversity in *Z. tritici* but also provides a basis for further search for virulence and adaptation of dependent genes, which should be explored in future experimental studies. For example, comparative analyses that may involve a large scale homology search in official databases or comparison with available transcriptome data could help to identify candidate genes that are maybe involved in virulence or the adaptation to the environment (Haueisen et al. 2017). By comparison between gene datasets of genes with positive diversifying and positive directional selection, this study finds a possible ABC transporter, a class of proteins associated with virulence and fungicide resistance, that shows signatures of both types of selection (De Waard 1997; Schoonbeek et al. 2002; Stergiopoulos et al. 2002). The link between ABC transport, virulence, and fungicide resistance underscores the importance of further investigating the putative ABC transporter for the further development of disease

management against *Z. tritici* (De Waard 1997; Schoonbeek et al. 2002; Stergiopoulos et al. 2002).

However, the found ABC transporter shows no strong homology with already known genes of this class. Therefore, further analyzes involving gene deletions and *in planta*, experiments should be performed in order to narrow down the potential function of the found ABC transporter.

To what extent the obtained results of the analyses of the natural selection depend on the observed differences of the genetic diversity must show further studies. For this, it will be essential to increase the population genomic data set by a further sampling of Fungal isolates of the species *Z. ardabiliae*, *Z. brevis*, *Z. pseudotritici,* and *Z. tritici*. Furthermore, further should check the found demographic differences for their plausibility using simulations.

In conclusion, the study presented here shows that the members of the *Zymoseptoria* species complex show substantial differences in the development of the effective population sizes (genetic diversity). Based on the results found, further comparative studies between natural and managed ecosystems could indicate which determinants (i.e., host density or host diversity) influence the effective population size of the pathogen to develop new agricultural methods to control the pathogen. Also, this study not only highlights the importance of species-specific effectors but also offers a large number of gene candidates that may be important for adaptation to the host or the rest of the environment.

7. References

Abel HJ, Duncavage E. 2015. Detection of structural DNA variation from next generation sequencing data: a review of informatic approaches. 206:432–440.

Aguileta G, Refrégier G, Yockteng R, Fournier E, Giraud T. 2009. Rapidly evolving genes in pathogens: Methods for detecting positive selection and examples among fungi, bacteria, viruses and protists. Infect. Genet. Evol. 9:656–670.

Alexander DH, Novembre J, Lange K. 2009. Fast model-based estimation of ancestry in unrelated individuals. Genome Res. 19:1655–1664.

Alizon S, Hurford A, Mideo N, Van Baalen M. 2009. Virulence evolution and the trade-off hypothesis: History, current state of affairs and the future. J. Evol. Biol.

Alizon S, de Roode JC, Michalakis Y. 2013. Multiple infections and the evolution of virulence. Ecol. Lett.

Alkan C, Eichler E, Coe BP. 2011. Genome structural variation discovery and genotyping. Nat. Rev. Genet. 257:2432–2437.

Andrews KR, Good JM, Miller MR, Luikart G, Hohenlohe PA. 2016. Harnessing the power of RADseq for ecological and evolutionary genomics. Natl. Rev. Genet. 17:81–92.

Angiuoli S V., Salzberg SL. 2011. Mugsy: Fast multiple alignment of closely related whole genomes. Bioinformatics 27:334–342.

Anisimova M, Bielawski J, Dunn K, Yang Z. 2007. Phylogenomic analysis of natural selection pressure in Streptococcus genomes. BMC Evol. Biol. 7.

Anisimova M, Bielawski JP, Yang Z. 2001. Accuracy and power of the likelihood ratio test in detecting adaptive molecular evolution. Mol. Biol. Evol.

Antipov D, Korobeynikov A, McLean JS, Pevzner PA. 2016. HybridSPAdes: An algorithm for hybrid assembly of short and long reads. Bioinformatics 32:1009–1015.

Baird NA, Etter PD, Atwood TS, Currey MC, Shiver AL, Lewis ZA, Selker EU, Cresko WA, Johnson EA. 2008. Rapid SNP discovery and genetic mapping using sequenced RAD markers. PLoS One 3:1–7.

Banke S, Mcdonald BA. 2005. Migration patterns among global populations of the pathogenic fungus Mycosphaerella graminicola. Mol. Ecol. 14:1881–1896.

Bankevich A, Nurk S, Antipov D, Gurevich AA, Dvorkin M, Kulikov AS, Lesin VM, Nikolenko SI, Pham S, Prjibelski AD, et al. 2012. SPAdes: A New Genome Assembly Algorithm and Its Applications to Single-Cell Sequencing. J. Comput. Biol. [Internet] 19:455–477. Available from: http://online.liebertpub.com/doi/abs/10.1089/cmb.2012.0021

Bao W, Kojima KK, Kohany O. 2015. Repbase Update, a database of repetitive elements in eukaryotic genomes. Mob. DNA 6.

Barnett DW, Garrison EK, Quinlan AR, Stŕmberg MP, Marth GT. 2011. Bamtools: A C++ API and toolkit for analyzing and managing BAM files. Bioinformatics 27:1691–1692.

Barton NH. 1995. Linkage and the limits to natural selection. Genetics [Internet] 140:821–841. Available from: http://www.genetics.org/content/140/2/821

Baxter L, Tripathy S, Ishaque N, Boot N, Cabral A, Kemen E, Thines M, Ah-Fong A, Anderson R, Badejoko W, et al. 2010. Signatures of adaptation to obligate biotrophy in the Hyaloperonospora arabidopsidis genome. Science (80-.). 330:1549–1551.

Bebber DP, Ramotowski MAT, Gurr SJ. 2013. Crop pests and pathogens move polewards in a warming world. Nat. Clim. Chang.

Benjamini Y, Hochberg Y. 1995. Controlling the false discovery rate: a practical and powerful approach to multiple testing. J. R. Stat. Soc. Ser. B.

Black IV WC, Baer CF, Antolin MF, DuTeau NM. 2002. Population Genomics: Genome-Wide Sampling of Insect Populations. Annu. Rev. Entomol. [Internet] 46:441–469. Available from: http://www.annualreviews.org/doi/10.1146/annurev.ento.46.1.441

Blanchette M. 2004. Aligning Multiple Genomic Sequences With the Threaded Blockset Aligner. Genome Res. [Internet] 14:708–715. Available from: http://www.genome.org/cgi/doi/10.1101/gr.1933104

Bokulich NA, Subramanian S, Faith JJ, Gevers D, Gordon I, Knight R, Mills DA, Caporaso JG. 2013. NIH Public Access. 10:57–59.

Bolger AM, Lohse M, Usadel B. 2014. Trimmomatic: A flexible trimmer for Illumina sequence data. Bioinformatics 30:2114–2120.

Bradbury PJ, Zhang Z, Kroon DE, Casstevens TM, Ramdoss Y, Buckler ES. 2007. TASSEL: Software for association mapping of complex traits in diverse samples. Bioinformatics 23:2633–2635.

Braker I. 2017. Karyotypic charachterization and population genomic analysis of the fungal grass pathogen *Zymoseptoria* ardabiliae. Univ. Kiel.

Brunner PC, Torriani SFF, Croll D, Stukenbrock EH, McDonald BA. 2013. Coevolution and life cycle specialization of plant cell wall degrading enzymes in a hemibiotrophic pathogen. Mol. Biol. Evol.
Burdon JJ. 1993. Pathogen Populations in. Office.
Burdon JJ, Chilvers GA. 2003. Host Density as a Factor in Plant Disease Ecology. Annu. Rev. Phytopathol.
Burdon JJ, Thrall PH. 2008. Pathogen evolution across the agro-ecological interface: implications for disease management. Evol. Appl.
Catchen J, Hohenlohe PA, Bassham S, Amores A, Cresko WA. 2013. Stacks: An analysis tool set for population genomics. Mol. Ecol. 22:3124–3140.
Catchen JM, Amores A, Hohenlohe P, Cresko W, Postlethwait JH. 2011. *Stacks*: Building and Genotyping Loci *De novo* From Short-Read Sequences. G3: Genes|Genomes|Genetics [Internet] 1:171–182. Available from: http://g3journal.org/lookup/doi/10.1534/g3.111.000240
Chakraborty M, Baldwin-Brown JG, Long AD, Emerson JJ. 2015. A practical guide to *de novo* genome assembly using long reads. bioRxiv [Internet]:29306. Available from: http://biorxiv.org/content/early/2015/10/16/029306.abstract
Charlesworth B, Charlesworth D, Barton NH. 2003. The effects of genetic and geographic structure on neutral variation. In: Annual Review of Ecology Evolution and Systematics.
Charlesworth D. 2006. Balancing selection and its effects on sequences in nearby genome regions. PLoS Genet.
Chen IC, Hill JK, Ohlemüller R, Roy DB, Thomas CD. 2011. Rapid range shifts of species associated with high levels of climate warming. Science (80-.).
Chen W, Wellings C, Chen X, Kang Z, Liu T. 2014. Wheat stripe (yellow) rust caused by Puccinia striiformis f. sp. tritici. Mol. Plant Pathol.
Chikhi L, Sousa V, Luisi P, Goossens B, Beaumont M. 2010. The Confounding Effects of Population Structure, Genetic Diversity and the Sampling Scheme on the Detection and Quantification of Population Size Changes. Genetics 3:983–995.
Chin C-S, Alexander DH, Marks P, Klammer AA, Drake J, Heiner C, Clum A, Copeland A, Huddleston J, Eichler EE, et al. 2013. Nonhybrid, finished microbial genome assemblies from long-read SMRT sequencing data. Nat. Methods [Internet] 10:563–569. Available from: http://www.nature.com/doifinder/10.1038/nmeth.2474
Chin C-S, Peluso P, Sedlazeck FJ, Nattestad M, Concepcion GT, Clum A, Dunn C, O'Malley R, Figueroa-Balderas R, Morales-Cruz A, et al. 2016. Phased diploid genome assembly with single-molecule real-time sequencing. Nat. Methods [Internet] 13:1050–1054. Available from: http://www.nature.com/doifinder/10.1038/nmeth.4035
Cissé OH, Ma L, Wei Huang D, Khil PP, Dekker JP, Kutty G, Bishop L, Liu Y, Deng X, Hauser PM, et al. 2018. Comparative population genomics analysis of the mammalian fungal pathogen Pneumocystis. MBio 9:1–17.
Clark BC. 1979. The evolution of genetic diversity. Proc. R. Soc. London B Biol. Sci. [Internet] 205:453–474. Available from: http://rspb.royalsocietypublishing.org/content/205/1161/453
Coleman J, Rounsley SD, Rodriguez-Carres M, Kuo A, Wasmann CC, Grimwood J, Schmutz J, Taga M, White GJ, Zhou S, et al. 2009. The Genome of Nectria haematococca: Contribution of Supernumerary Chromosomes to Gene Expansion. PLoS Genet 5:e1000618.
Couch BC, Fudal I, Lebrun MH, Tharreau D, Valent B, Van Kim P, Nottéghem JL, Kohn LM. 2005. Origins of host-specific populations of the blast pathogen Magnaporthe oryzae in crop domestication with subsequent expansion of pandemic clones on rice and weeds of rice. Genetics 170:613–630.
Croll D, Lendenmann MH, Stewart E, McDonald BA. 2015. The impact of recombination hotspots on genome evolution of a fungal plant pathogen. Genetics 201:1213–1228.
Croll D, McDonald BA. 2012. The accessory genome as a cradle for adaptive evolution in pathogens. PLoS Pathog. 8:8–10.
Croll D, McDonald BA. 2017. The genetic basis of local adaptation for pathogenic fungi in agricultural ecosystems. Mol. Ecol.
Crouch J, O'Connell R, Gan P, Buiate E, Torres MF, Beirn L, Shirasu K, Vaillancourt L. 2014. The genomics of colletotrichum. In: Genomics of Plant-Associated Fungi: Monocot Pathogens. p. 69–102.
Crous PW, Zala M, Quaedvlieg W, Stukenbrock EH, Javan-Nikhah M, McDonald BA. 2012. *Zymoseptoria* ardabiliae and *Z. pseudotritici* , two progenitor species of the septoria tritici leaf blotch fungus *Z. tritici* (synonym: Mycosphaerella graminicola) . Mycologia [Internet] 104:1397–1407. Available from: http://www.ncbi.nlm.nih.gov/pubmed/22675045
Cuomo CA, Güldener U, Xu J-R, Trail F, Turgeon B. G, Di Pietro A, Walton JD, Ma L-J, Baker SE, Rep M, et al. 2007. The Fusarium graminearum genome reveals a link between localized polymorphism and pathogen specialization. Science (80-.). [Internet] 317:1400–1402. Available from:

http://www.ncbi.nlm.nih.gov/pubmed/17823352
Danecek P, Auton A, Abecasis G, Albers CA, Banks E, DePristo MA, Handsaker RE, Lunter G, Marth GT, Sherry ST, et al. 2011. The variant call format and VCFtools. Bioinformatics 27:2156–2158.
Dangl JL, Jones JDG. 2001. Plant pathogens and integrated defence responses to infection. Nature [Internet] 411:826–833. Available from: http://www.nature.com/doifinder/10.1038/35081161
Darling AE, Mau B, Perna NT. 2010. Progressivemauve: Multiple genome alignment with gene gain, loss and rearrangement. PLoS One 5.
Davey JW, Cezard T, Fuentes-Utrilla P, Eland C, Gharbi K, Blaxter ML. 2013. Special features of RAD Sequencing data: Implications for genotyping. Mol. Ecol. 22:3151–3164.
Day J, Gietz DD, Rampitsch C. 2015. Proteome changes induced by Pyrenophora tritici-repentis ToxA in both insensitive and sensitive wheat indicate senescence-like signaling. Proteome Sci.
Dean R, Van Kan JALL, Pretorius ZA, Hammond-Kosack KE, Di Pietro A, Spanu PD, Rudd JJ, Dickman M, Kahmann R, Ellis J, et al. 2012. The Top 10 fungal pathogens in molecular plant pathology. Mol. Plant Pathol. 13:414–430.
Dean RA, Talbot NJ, Ebbole DJ, Farman ML, Mitchell TK, Orbach MJ, Thon M, Kulkarni R, Xu JR, Pan H, et al. 2005. The genome sequence of the rice blast fungus Magnaporthe grisea. Nature 434:980–986.
Desjardins CA, Giamberardino C, Sykes SM, Yu C-HH, Tenor JL, Chen Y, Yang T, Jones AM, Sun S, Haverkamp MR, et al. 2017. Population genomics and the evolution of virulence in the fungal pathogen Cryptococcus neoformans. Genome Res. 27:1207–1219.
Dillard HR. 2019. Global food and nutrition security: from challenges to solutions. Food Secur.:249–252.
Dodds PN. 2010. Genome evolution in plant pathogens. Science (80-.).
Dodds PN, Rathjen JP. 2010. Plant immunity: towards an integrated view of plant–pathogen interactions. Nat. Rev. Genet. [Internet] 11:539–548. Available from: http://www.nature.com/doifinder/10.1038/nrg2812
Doebley JF, Gaut BS, Smith BD. 2006. The Molecular Genetics of Crop Domestication. :1309–1321.
van der Does HC, Fokkens L, Yang A, Schmidt SM, Langereis L, Lukasiewicz JM, Hughes TR, Rep M. 2016. Transcription Factors Encoded on Core and Accessory Chromosomes of Fusarium oxysporum Induce Expression of Effector Genes. PLoS Genet. 12.
van der Does HC, Rep M. 2007. Virulence Genes and the Evolution of Host Specificity in Plant-Pathogenic Fungi. Mol. Plant-Microbe Interact.
Dolgin ES, Charlesworth B. 2006. The fate of transposable elements in asexual populations. Genetics 174:817–827.
Duplessis S, Cuomo CA, Lin Y-C, Aerts A, Tisserant E, Veneault-Fourrey C, Joly DL, Hacquard S, Amselem J, Cantarel BL, et al. 2011. Obligate biotrophy features unraveled by the genomic analysis of rust fungi. Proc. Natl. Acad. Sci.
Dutheil JY, Gaillard S, Stukenbrock EH. 2014. MafFilter: a highly flexible and extensible multiple genome alignment files processor. BMC Genomics [Internet] 15:53. Available from: http://bmcgenomics.biomedcentral.com/articles/10.1186/1471-2164-15-53
Dutheil JY, Mannhaupt G, Schweizer G, M K Sieber C, Münsterkötter M, Güldener U, Schirawski J, Kahmann R. 2016. A Tale of Genome Compartmentalization: The Evolution of Virulence Clusters in Smut Fungi. Genome Biol. Evol. 8:681–704.
Edgar RC, Flyvbjerg H. 2015. Error filtering, pair assembly and error correction for next-generation sequencing reads. Bioinformatics 31:3476–3482.
Eid J, Fehr A, Gray J, Luong K, Lyle J, Otto G, Peluso P, Rank D, Baybayan P, Bettman B, et al. 2009. Real-Time DNA Sequencing from Single Polymerase Molecules. Science (80-.). [Internet] 323:133–138. Available from: http://www.sciencemag.org/cgi/doi/10.1126/science.1162986
Ekblom R, Wolf JBW. 2014. A field guide to whole-genome sequencing, assembly and annotation. Evol. Appl. 7:1026–1042.
Ellegren H, Galtier N. 2016. Determinants of genetic diversity. Nat. Rev. Genet. [Internet] 17:422–433. Available from: http://dx.doi.org/10.1038/nrg.2016.58
Elshire RJ, Glaubitz JC, Sun Q, Poland JA, Kawamoto K, Buckler ES, Mitchell SE. 2011. A robust, simple genotyping-by-sequencing (GBS) approach for high diversity species. PLoS One 6:1–10.
Emanuelsson O, Nielsen H, Brunak S, von Heijne G. 2000. Predicting subcellular localization of proteins based on their N-terminal amino acid sequence. J. Mol. Biol.
Eschenbrenner CJ. 2013. The fungal grass pathogen *Zymoseptoria* brevis, a new model in the *Zymoseptoria* species complex. Univ. Marbg.
Evenson RE, Gollin D. 2003. Assessing the impact of the Green Revolution, 1960 to 2000. Science (80-.).
Eyal Z, Amiri Z, Wahl I. 1973. Physiological Specialization of Septoria tritici. Phytopathology 63:1087–1091.
Eyal Z, Scharen A, Huffman M, Prescott J. 1985. Global insights into virulence frequencies of

Mycosphaerella graminicola. Phytopathology [Internet] 75:1456–1462. Available from: http://www.apsnet.org/publications/phytopathology/backissues/Documents/1985Articles/Phyto75n12_1456.PDF
Eyre-Walker A. 2006. The genomic rate of adaptive evolution. Trends Ecol. Evol.
Faino L, Seidl MF, Shi-Kunne X, Pauper M, Van Den Berg GCM, Wittenberg AHJ, Thomma BPHJ. 2016. Transposons passively and actively contribute to evolution of the two-speed genome of a fungal pathogen. Genome Res. 26:1091–1100.
FAO. 2018. Statistical Pocketbook 2018. World food and agriculture. Rome Available from: licence: CC BY-NC-SA 3.0 IGO
Fawcett JA, Iida T, Takuno S, Sugino RP, Kado T, Kugou K, Mura S, Kobayashi T, Ohta K, Nakayama JI, et al. 2014. Population genomics of the fission yeast Schizosaccharomyces pombe. PLoS One 9.
Feldbrügge M, Kämper J, Steinberg G, Kahmann R. 2004. Regulation of mating and pathogenic development in Ustilago maydis. Curr. Opin. Microbiol. 7:666–672.
Felsenstein J. 1989. PHYLIP -- Phylogeny Inference Package (Version 3.2). Cladistics.
Felsenstein J. 2005. PHYLIP (Phylogeny Inference Package). Distrib. by author Dep. Genome Sci. Univ. Washingt. Seattle.
Fierro F, Martín JF. 1999. Molecular Mechanisms of Chromosomal Rearrangement in Fungi. Crit. Rev. Microbiol. [Internet] 25:1–17. Available from: http://www.tandfonline.com/doi/full/10.1080/10408419991299185
Fisher MC, Henk D a, Briggs CJ, Brownstein JS, Madoff LC, McCraw SL, Gurr SJ. 2012. Emerging fungal threats to animal, plant and ecosystem health. Nature [Internet] 484:186–194. Available from: http://www.ncbi.nlm.nih.gov/pubmed/22498624
Flicek P, Birney E. 2010. Sense from sequence reads: methods for alignment and assembly. Nat. Methods [Internet] 7:479–479. Available from: http://www.nature.com/doifinder/10.1038/nmeth0610-479b
Flint-Garcia SA, Thornsberry JM, Buckler ES. 2003. Structure of Linkage Disequilibrium in Plants. Annu. Rev. Plant Biol.
Foley JA, Ramankutty N, Brauman KA, Cassidy ES, Gerber JS, Johnston M, Mueller ND, O'Connell C, Ray DK, West PC, et al. 2011. Solutions for a cultivated planet. Nature.
Fones H, Gurr S. 2015. The impact of Septoria tritici Blotch disease on wheat: An EU perspective. Fungal Genet. Biol.
Forouzan E, Maleki MSM, Karkhane AA, Yakhchali B. 2017. Evaluation of nine popular *de novo* assemblers in microbial genome assembly. J. Microbiol. Methods [Internet] 143:32–37. Available from: http://dx.doi.org/10.1016/j.mimet.2017.09.008
Fouché S, Plissonneau C, McDonald BA, Croll D. 2018. Meiosis leads to pervasive copy-number variation and distorted inheritance of accessory chromosomes of the wheat pathogen *Zymoseptoria* tritici. Genome Biol. Evol.
Friesen TL, Stukenbrock EH, Liu Z, Meinhardt S, Ling H, Faris JD, Rasmussen JB, Solomon PS, McDonald BA, Oliver RP. 2006. Emergence of a new disease as a result of interspecific virulence gene transfer. Nat. Genet. [Internet] 38:953–956. Available from: http://www.ncbi.nlm.nih.gov/pubmed/16832356
Frith MC, Hamada M, Horton P. 2010. Parameters for accurate genome alignment. BMC Bioinformatics 11.
Fulnečková J, Ševčíková T, Fajkus J, Lukešová A, Lukeš M, Vlček Č, Lang BF, Kim E, Eliáš M, Sýkorová E. 2013. A broad phylogenetic survey unveils the diversity and evolution of telomeres in eukaryotes. Genome Biol. Evol. 5:468–483.
Fumagalli M. 2013. Assessing the effect of sequencing depth and sample size in population genetics inferences. PLoS One.
Gabriel W, Lynch M, Burger R. 1993. Muller' s Ratchet and Mutational Meltdowns. Evolution (N. Y).
Galazka JM, Freitag M. 2014. Variability of chromosome structure in pathogenic fungi-of "ends and odds." Curr. Opin. Microbiol. 20:19–26.
Gange AC, Gange EG, Mohammad AB, Boddy L. 2011. Host shifts in fungi caused by climate change? Fungal Ecol.
García-Alcalde F, Okonechnikov K, Carbonell J, Cruz LM, Götz S, Tarazona S, Dopazo J, Meyer TF, Conesa A. 2012. Qualimap: Evaluating next-generation sequencing alignment data. Bioinformatics 28:2678–2679.
Gel B, Díez-Villanueva A, Serra E, Buschbeck M, Peinado MA, Malinverni R. 2016. regioneR: an R/Bioconductor package for the association analysis of genomic regions based on permutation tests. Bioinformatics [Internet] 32:289–291. Available from: http://bioinformatics.oxfordjournals.org/content/32/2/289%5Cnhttp://bioinformatics.oxfordjournals.org/content/32/2/289.full.pdf%5Cnhttp://bioinformatics.oxfordjournals.org/content/32/2/289.long%

5Cnhttp://www.ncbi.nlm.nih.gov/pubmed/26424858
Gilbert G. 2002. Evolutionary Ecology of Plant Diseases in Natural Ecosystems. Annu. Rev. Phytopathol.
Gingeras TR, Milazzo JP, Sciaky D, Roberts RJ. 1979. Computer programs for the assembly of DNA sequences. Nucleic Acids Res. 7:529–543.
Van Ginkel M, Scharen AL. 1987. Generation mean analysis and heritabilities of resistance to Septoria tritici in durum wheat. Phytopathology 77:1629–1633.
Giraldo MC, Valent B. 2013. Filamentous plant pathogen effectors in action. Nat. Rev. Microbiol. [Internet] 11:800–814. Available from: http://www.nature.com/doifinder/10.1038/nrmicro3119
Gladieux P, Ravel S, Rieux A, Cros-Arteil S, Adreit H, Milazzo J, Thierry M, Fournier E, Terauchi R, Tharreau D. 2018. Coexistence of multiple endemic and pandemic lineages of the rice blast pathogen. MBio 9:1–18.
Godfray HCJ, Beddington JR, Crute IR, Haddad L, Lawrence D, Muir JF, Pretty J, Robinson S, Thomas SM, Toulmin C. 2010. Food security: The challenge of feeding 9 billion people. Science (80-.).
Goldman N, Yangf Z. 1994. A codon-based model of nucleotide substitution for protein-coding DNA sequences. Mol. Biol. Evol. [Internet]. Available from: https://academic.oup.com/mbe/article/11/5/725/1008711/A-codonbased-model-of-nucleotide-substitution-for
Goodwin SB, Cohen BA, Deahl KL, Fry WE. 1994. Migration from northern Mexico as the probable cause of recent genetic changes in populations of Phytophthora infestans in the United States and Canada. Phytopathology.
Goodwin SB, M'Barek S Ben, Dhillon B, Wittenberg AHJ, Crane CF, Hane JK, Foster AJ, van der Lee TAJ, Grimwood J, Aerts A, et al. 2011. Finished genome of the fungal wheat pathogen Mycosphaerella graminicola reveals dispensome structure, chromosome plasticity, and stealth pathogenesis. PLoS Genet. 7.
Goss EM, Cardenas ME, Myers K, Forbes GA, Fry WE, Restrepo S, Grünwald NJ. 2011. The plant pathogen phytophthora andina emerged via hybridization of an unknown phytophthora species and the irish potato famine pathogen, P. infestans. PLoS One.
Götz S, García-Gómez JM, Terol J, Williams TD, Nagaraj SH, Nueda MJ, Robles M, Talón M, Dopazo J, Conesa A. 2008. High-throughput functional annotation and data mining with the Blast2GO suite. Nucleic Acids Res.
Grandaubert J, Bhattacharyya A, Stukenbrock EH. 2015. RNA-seq-Based Gene Annotation and Comparative Genomics of Four Fungal Grass Pathogens in the Genus *Zymoseptoria* Identify Novel Orphan Genes and Species-Specific Invasions of Transposable Elements. G3: Genes|Genomes|Genetics [Internet] 5:1323–1333. Available from: http://g3journal.org/lookup/doi/10.1534/g3.115.017731
Grandaubert J, Dutheil JY, Stukenbrock EH. 2017. The genomic rate of adaptation in the fungal wheat pathogen *Zymoseptoria* tritici. Doi.Org [Internet]:176727. Available from: https://www.biorxiv.org/content/early/2017/08/15/176727
Grünwald NJ, Everhart SE, Knaus BJ, Kamvar ZN. 2017. Best Practices for Population Genetic Analyses. Phytopathology [Internet] 107:1000–1010. Available from: http://apsjournals.apsnet.org/doi/10.1094/PHYTO-12-16-0425-RVW
Grünwald NJ, McDonald BA, Milgroom MG. 2016. Population Genomics of Fungal and Oomycete Pathogens. Annu. Rev. Phytopathol. [Internet] 54:323–346. Available from: http://www.annualreviews.org/doi/10.1146/annurev-phyto-080614-115913
Guindon S, Dufayard JF, Lefort V, Anisimova M, Hordijk W, Gascuel O. 2010. New algorithms and methods to estimate maximum-likelihood phylogenies: Assessing the performance of PhyML 3.0. Syst. Biol.
Gurevich A, Saveliev V, Vyahhi N, Tesler G. 2013. QUAST: Quality assessment tool for genome assemblies. Bioinformatics 29:1072–1075.
Gurr GM, Wratten SD, Michael Luna J. 2003. Multi-function agricultural biodiversity: Pest management and other benefits. Basic Appl. Ecol.
Gutiérrez-Alonso O, Hawkins NJ, Cools HJ, Shaw MW, Fraaije BA. 2017. Dose-dependent selection drives lineage replacement during the experimental evolution of SDHI fungicide resistance in *Zymoseptoria* tritici. Evol. Appl.
Haas BJ, Delcher AL, Wortman JR, Salzberg SL. 2004. DAGchainer: A tool for mining segmental genome duplications and synteny. Bioinformatics 20:3643–3646.
Habig M, Kema G, Holtgrewe Stukenbrock E. 2018. Meiotic drive of female-inherited supernumerary chromosomes in a pathogenic fungus. bioRxiv [Internet]. Available from: https://www.biorxiv.org/content/early/2018/08/10/389650
Habig M, Quade JJJ, Stukenbrock EH. 2017. Forward genetics approach reveals host genotype-dependent importance of accessory chromosomes in the fungal wheat pathogen *Zymoseptoria*

tritici. MBio 8:1–16.
Hacquard S, Kracher B, Hiruma K, Münch PC, Garrido-Oter R, Thon MR, Weimann A, Damm U, Dallery JF, Hainaut M, et al. 2016. Survival trade-offs in plant roots during colonization by closely related beneficial and pathogenic fungi. Nat. Commun. 7.
Hacquard S, Kracher B, Maekawa T, Vernaldi S, Schulze-Lefert P, Ver Loren van Themaat E. 2013. Mosaic genome structure of the barley powdery mildew pathogen and conservation of transcriptional programs in divergent hosts. Proc. Natl. Acad. Sci. [Internet] 110:E2219–E2228. Available from: http://www.pnas.org/cgi/doi/10.1073/pnas.1306807110
Haddrill PR, Thornton KR, Charlesworth B, Andolfatto P. 2005. Multilocus patterns of nucleotide variability and the demographic and selection history of Drosophila melanogaster populations. Genome Res.
Hamilton WD. 1980. Sex versus Non-Sex versus Parasite. Oikos.
Han Y, Liu X, Benny U, Corby Kistler H, VanEtten HD. 2001. Genes determining pathogenicity to pea are clustered on a supernumerary chromosome in the fungal plant pathogen Nectria haematococca. Plant J.
Hane JK, Rouxel T, Howlett BJ, Kema GHHJ, Goodwin SB, Oliver RP. 2011. A novel mode of chromosomal evolution peculiar to filamentous Ascomycete fungi. Genome Biol. [Internet] 12:R45. Available from: http://genomebiology.biomedcentral.com/articles/10.1186/gb-2011-12-5-r45
Hardison RC. 2003. Comparative Genomics. PLOS Biol. [Internet] 1. Available from: https://doi.org/10.1371/journal.pbio.0000058
Hartl DL, Clark AG. 2007. Principles of Population Genetics. 4th ed. (Sinauer AD, editor.). Sunderland: Sinauer Associates
Hartmann FE, Sánchez-Vallet A, McDonald BA, Croll D. 2017. A fungal wheat pathogen evolved host specialization by extensive chromosomal rearrangements. ISME J.
Hatta R, Ito K, Hosaki Y, Tanaka T, Tanaka A, Yamamoto M, Akimitsu K, Tsuge T. 2002. A conditionally dispensable chromosome controls host-specific pathogenicity in the fungal plant pathogen Alternaria alternata. Genetics.
Haueisen J, Moeller M, Eschenbrenner CJ, Grandaubert J, Seybold H, Adamiak H, Stukenbrock EH. 2017. Extremely flexible infection programs in a fungal plant pathogen. bioRxiv.
Haussler D, O'Brien SJ, Ryder OA, Keith Barker F, Clamp M, Crawford AJ, Hanner R, Hanotte O, Johnson WE, McGuire JA, et al. 2009. Genome 10K: A proposal to obtain whole-genome sequence for 10000 vertebrate species. J. Hered. 100:659–674.
Hayes BJ, Visscher PM, McPartlan HC, Goddard ME. 2003. Novel multilocus measure of linkage disequilibrium to estimate past effective population size. Genome Res.
Herschleb J, Ananiev G, Schwartz DC. 2007. Pulsed-field gel electrophoresis. Nat. Protoc. 2:677–684.
Hill WG. 1981. Estimation of effective population size from data on linkage disequilibrium. Genet. Res.
Hill WG, Robertson A. 1966. The effect of linkage on limits to artificial selection. Genet. Res. 8:269–294.
Hill WG, Robertson A. 1968. Linkage disequilibrium in finite populations. Theor. Appl. Genet.
Hohenlohe PA, Day MD, Amish SJ, Miller MR, Kamps-Hughes N, Boyer MC, Muhlfeld CC, Allendorf FW, Johnson EA, Luikart G. 2013. Genomic patterns of introgression in rainbow and westslope cutthroat trout illuminated by overlapping paired-end RAD sequencing. Mol. Ecol. 22:3002–3013.
Houben A. 2017. B Chromosomes – A Matter of Chromosome Drive. Front. Plant Sci.
Hovmøller MS, Sørensen CK, Walter S, Justesen AF. 2011. Diversity of Puccinia striiformis on Cereals and Grasses . Annu. Rev. Phytopathol.
Hughes AL. 2007. Looking for Darwin in all the wrong places: The misguided quest for positive selection at the nucleotide sequence level. Heredity (Edinb).
Igarashi S, Utiamada CM, Igarashi LC, Kazuma AH, Lopes RS. 1986. Pyricularia in wheat. 1. Occurrence of Pyricularia sp. Paran state. Fitopatol. Bras 11:351–352.
Islam MT, Croll D, Gladieux P, Soanes DM, Persoons A, Bhattacharjee P, Hossain MS, Gupta DR, Rahman MM, Mahboob MG, et al. 2016. Emergence of wheat blast in Bangladesh was caused by a South American lineage of Magnaporthe oryzae. BMC Biol. [Internet] 14:1–11. Available from: http://dx.doi.org/10.1186/s12915-016-0309-7
Jackson AP, Gamble JA, Yeomans T, Moran GP, Saunders D, Harris D, Aslett M, Barrell JF, Butler G, Citiulo F, et al. 2009. Comparative genomics of the fungal pathogens Candida dubliniensis and Candida albicans. Genome Res.
Jaenike J. 1978. An hypothesis to account for the maintenance of sex within populations. Evol. Theory.
Javelle M, Vernoud V, Rogowsky PM, Ingram GC. 2011. Epidermis: The formation and functions of a fundamental plant tissue. New Phytol. 189:17–39.
Jayakar SD. 1970. A mathematical model for interaction of gene frequencies in a parasite and its host. Theor. Popul. Biol.
Jones JDG, Dangl JL. 2006. The plant immune system. Nature [Internet] 444:323–329. Available from: http://www.nature.com/doifinder/10.1038/nature05286

Jones L, Riaz S, Morales-Cruz A, Amrine KCH, McGuire B, Gubler WD, Walker MA, Cantu D. 2014. Adaptive genomic structural variation in the grape powdery mildew pathogen, Erysiphe necator. BMC Genomics.

Jones N. 2012. B chromosomes in plants. Plant Biosyst.

De Jonge R, Bolton MD, Kombrink A, Van Den Berg GCM, Yadeta KA, Thomma BPHJ. 2013. Extensive chromosomal reshuffling drives evolution of virulence in an asexual pathogen. Genome Res. 23:1271–1282.

Kaku H, Nishizawa Y, Ishii-Minami N, Akimoto-Tomiyama C, Dohmae N, Takio K, Minami E, Shibuya N. 2006. Plant cells recognize chitin fragments for defense signaling through a plasma membrane receptor. Proc. Natl. Acad. Sci.

Karimi V, Karami E, Keshavarz M. 2018. Climate change and agriculture: Impacts and adaptive responses in Iran. J. Integr. Agric.

Katagiri F, Tsuda K. 2010. Understanding the Plant Immune System. Mol. Plant-Microbe Interact. [Internet] 23:1531–1536. Available from: http://apsjournals.apsnet.org/doi/10.1094/MPMI-04-10-0099

Katoh K, Asimenos G, Toh H. 2009. Multiple alignment of DNA sequences with MAFFT. Methods Mol. Biol. 537:39–64.

Keller SM, McDermott JM, Pettway RE, Wolfe MS, McDonald BA. 1997. Gene Flow and Sexual Reproduction in the Wheat Glume Blotch Pathogen *Phaeosphaeria nodorum* (Anamorph *Stagonospora nodorum*). Phytopathology.

Kelley CP, Mohtadi S, Cane MA, Seager R, Kushnir Y. 2015. Climate change in the Fertile Crescent and implications of the recent Syrian drought. Proc. Natl. Acad. Sci.

Kema GHJ, Yu D, Shaw MW, Baayen RP, Rijkenberg FHJ. 1996. Histology of the pathogenesis of Mycosphaerella graminicola in wheat. Phytopathology.

Kim Y, Stephan W. 2002. Detecting a local signature of genetic hitchhiking along a recombining chromosome. Genetics 160:765–777.

Kimura M. 1968. Evolutionary rate at the molecular level. Nature.

Kimura M. 1969. The number of heterozygous nucleotide sites maintained in a finite population due to steady flux of mutations. Genetics.

Kimura M. 1971. Theoretical foundation of population genetics at the molecular level. Theor. Popul. Biol. [Internet] 2:174–208. Available from: http://linkinghub.elsevier.com/retrieve/pii/0040580971900141

Kimura M. 1977. Preponderance of synonymous changes as evidence for the neutral theory of molecular evolution [33]. Nature.

King R, Brown NA, Urban M, Hammond-Kosack KE. 2018. Inter-genome comparison of the Quorn fungus Fusarium venenatum and the closely related plant infecting pathogen Fusarium graminearum. BMC Genomics.

King R, Urban M, Hammond-Kosack MCU, Hassani-Pak K, Hammond-Kosack KE. 2015. The completed genome sequence of the pathogenic ascomycete fungus Fusarium graminearum. BMC Genomics [Internet] 16:1–21. Available from: http://dx.doi.org/10.1186/s12864-015-1756-1

Kliman R, Sheehy B, Schultz J. 2008. Genetic Drift and Effective Population Size. Nat. Educ.

Kobmoo N, Wichadakul D, Arnamnart N, Rodríguez De La Vega RC, Luangsa-ard JJ, Giraud T. 2018. A genome scan of diversifying selection in Ophiocordyceps zombie-ant fungi suggests a role for enterotoxins in co-evolution and host specificity. Mol. Ecol.

Koren S, Walenz BP, Berlin K, Miller JR, Bergman NH, Phillippy AM. 2016. Canu : scalable and accurate long- - - read assembly via adaptive k - - - mer weighting and repeat separation. :1–35.

Kreitman M. 2000. Methods to detect selection in populations with applications to the human. Annu. Rev. Genomics Hum. Genet.

Krishnan P, Ma X, McDonald BA, Brunner PC. 2018. Widespread signatures of selection for secreted peptidases in a fungal plant pathogen. BMC Evol. Biol.

Krzywinski M, Schein J, Birol I, Connors J, Gascoyne R, Horsman D, Jones SJ, Marra MA. 2009. Circos: An information aesthetic for comparative genomics. Genome Res.

Kurtz S, Phillippy A, Delcher AL, Smoot M, Shumway M, Antonescu C, Salzberg SL. 2004. Versatile and open software for comparing large genomes. Genome Biol. [Internet] 5:R12. Available from: http://genomebiology.com/2004/5/2/R12

Langmead B, Trapnell C, Pop M, Salzberg S. 2009. Ultrafast and memory-efficient alignment of short DNA sequences to the human genome. Genome Biol. [Internet] 10:R25. Available from: papers://1bc19a7c-6e6a-4594-831a-36d8c340e116/Paper/p2438

Layer RM, Chiang C, Quinlan AR, Hall IM. 2014. LUMPY: A probabilistic framework for structural variant discovery. Genome Biol.

Lenoir J, Svenning JC. 2015. Climate-related range shifts - a global multidimensional synthesis and new research directions. Ecography (Cop.).

Letunic I, Bork P. 2018. 20 years of the SMART protein domain annotation resource. Nucleic Acids Res.
Letunic I, Doerks T, Bork P. 2015. SMART: Recent updates, new developments and status in 2015. Nucleic Acids Res.
Li H. 2011. A statistical framework for SNP calling, mutation discovery, association mapping and population genetical parameter estimation from sequencing data. Bioinformatics 27:2987–2993.
Li H, Durbin R. 2009. Fast and accurate short read alignment with Burrows-Wheeler transform. Bioinformatics 25:1754–1760.
Li H, Durbin R. 2010. Fast and accurate long-read alignment with Burrows-Wheeler transform. Bioinformatics 26:589–595.
Li H, Handsaker B, Wysoker A, Fennell T, Ruan J, Homer N, Marth G, Abecasis G, Durbin R. 2009. The Sequence Alignment/Map format and SAMtools. Bioinformatics 25:2078–2079.
Li R, Li Y, Fang X, Res G, Li R, Li Y, Fang X, Yang H, Wang J, Kristiansen K, et al. 2009. SNP detection for massively parallel whole-genome resequencing SNP detection for massively parallel whole-genome resequencing. :1124–1132.
Lindberg MR, Hall IM, Quinlan AR. 2015. Population-based structural variation discovery with Hydra-Multi. Bioinformatics 31:1286–1289.
Linde CC, Zhan J, McDonald BA. 2002. Population structure of Mycosphaerella graminicola: From lesions to continents. Phytopathology [Internet] 92:946–955. Available from: http://apsjournals.apsnet.org/doi/abs/10.1094/PHYTO.2002.92.9.946
Lindenbaum P. 2015. JVarkit: java-based utilities for Bioinformatics. Figshare [Internet]:2–5. Available from: https://dx.doi.org/10.6084/m9.figshare.1425030.v1
Liu X, Han S, Wang Z, Gelernter J, Yang BZ. 2013. Variant Callers for Next-Generation Sequencing Data: A Comparison Study. PLoS One 8:1–11.
Liu X, Xiao G, Hu X, Zhan S, Su Y, Zhang X, St. Leger RJ, Wang C, Zheng P, Shang Y. 2014. Trajectory and genomic determinants of fungal-pathogen speciation and host adaptation. Proc. Natl. Acad. Sci.
Luikart G. 2014. Recent novel approaches for population genomics data analysis. :1661–1667.
Luo R, Liu B, Xie Y, Li Z, Huang W, Yuan J, Wang J. 2012. SOAPdenovo2: an empirically improved memory-efficient short-read *de novo* assembler. Gigascience 1:18.
Lynch M. 2003. The Origins of Genome Complexity. Science (80-.). [Internet] 302:1401–1404. Available from: http://www.sciencemag.org/cgi/doi/10.1126/science.1089370
Ma L-JJ, Van Der Does HC, Borkovich KA, Coleman JJ, Daboussi M-JJ, Di Pietro A, Dufresne M, Freitag M, Grabherr M, Henrissat B, et al. 2010. Comparative genomics reveals mobile pathogenicity chromosomes in Fusarium. Nature [Internet] 464:367–373. Available from: http://www.nature.com/doifinder/10.1038/nature08850
Maciel JLN, Ceresini PC, Castroagudin VL, Zala M, Kema GHJ, McDonald BA. 2014. Population Structure and Pathotype Diversity of the Wheat Blast Pathogen *Magnaporthe oryzae* 25 Years After Its Emergence in Brazil. Phytopathology.
Malaker PK, Barma NCD, Tewari TP, Collis WJ, Duveiller E, Singh PK, Joshi AK, Singh RP, Braun H-J, Peterson GL, et al. 2016. First report of wheat blast caused by Magnaporthe oryzae pathotype triticum in Bangladesh. Plant Dis.
Malaspinas AS, Westaway MC, Muller C, Sousa VC, Lao O, Alves I, Bergström A, Athanasiadis G, Cheng JY, Crawford JE, et al. 2016. A genomic history of Aboriginal Australia. Nature.
Manning VA. 2005. Localization of Ptr ToxA Produced by Pyrenophora tritici-repentis Reveals Protein Import into Wheat Mesophyll Cells. PLANT CELL ONLINE.
Marais G, Charlesworth B. 2003. Genome evolution: Recombination speeds up adaptive evolution. Curr. Biol. 13.
Marschall T, Marz M, Abeel T, Dijkstra L, Dutilh BE, Ghaffaari A, Kersey P, Kloosterman WP, Mäkinen V, Novak AM, et al. 2018. Computational pan-genomics: Status, promises and challenges. Brief. Bioinform. 19:118–135.
Martin F, Kamoun S. 2011. Effectors in Plant-Microbe Interactions.
Maruyama T, Fuerst PA. 1985. Population bottlenecks and nonequilibrium models in population genetics. II. Number of alleles in a small population that was formed by a recent bottleneck. Genetics.
Maynard Smith J, Haigh J. 2008. The hitch-hiking effect of a favourable gene. Genet. Res. (Camb).
Mayr E. 1954. Change of genetic environment and evolution. In: Huxley J, Hardy A, Ford E, editors. Evolution as a Process. London: Allen & Unwin. p. 157–180.
McDonald B a, Zhan J, Burdon JJ. 1999. Genetic Structure of Rhynchosporium secalis in Australia. Phytopathology [Internet] 89:639–645. Available from: http://apsjournals.apsnet.org/
McDonald BA. 2014. Using dynamic diversity to achieve durable disease resistance in agricultural ecosystems. Trop. Plant Pathol.
McDonald BA, Linde C. 2002. Pathogen population genetics, evolutionary potential, and durable

resistance. Annu. Rev. Phytopathol. 40:349–379.
McDonald BA, Martinez JP. 1991. Chromosome length polymorphisms in a Septoria tritici population. Curr. Genet.
McDonald BA, Stukenbrock EH. 2016. Rapid emergence of pathogens in agro-ecosystems: global threats to agricultural sustainability and food security. Philos. Trans. R. Soc. B Biol. Sci. [Internet] 371:20160026. Available from: http://rstb.royalsocietypublishing.org/lookup/doi/10.1098/rstb.2016.0026
McDonald JH, Kreitman M. 1991. Adaptive protein evolution at the Adh locus in Drosophila. Nature [Internet] 351:652–654. Available from: http://www.nature.com/doifinder/10.1038/351652a0
McDonald MC, Oliver RP, Friesen TL, Brunner PC, McDonald BA. 2013. Global diversity and distribution of three necrotrophic effectors in Phaeosphaeria nodorum and related species. New Phytol.
McKenna A, Hanna M, Banks E, Sivachenko A, Cibulskis K, Kernytsky A, Garimella K, Altshuler D, Gabriel S, Daly M, et al. 2010. The Genome Analysis Toolkit: A MapReduce framework for analyzing next-generation DNA sequencing data. Genome Res. [Internet] 20:1297–1303. Available from: http://genome.cshlp.org/cgi/doi/10.1101/gr.107524.110
Medini D, Donati C, Tettelin H, Masignani V, Rappuoli R. 2005. The microbial pan-genome. Curr. Opin. Genet. Dev. 15:589–594.
Medvedev P, Fiume M, Dzamba M, Smith T, Brudno M. 2010. Detecting copy number variation with mated short reads. Genome Res.
De Meeûs T, Prugnolle F, Agnew P. 2007. Asexual reproduction: Genetics and evolutionary aspects. Cell. Mol. Life Sci.
Mehrabi R, Mirzadi Gohari A, Kema GHJ. 2017. Karyotype Variability in Plant-Pathogenic Fungi. Annu. Rev. Phytopathol. [Internet] 55:483–503. Available from: http://www.annualreviews.org/doi/10.1146/annurev-phyto-080615-095928
Mehrabi R, Taga M, Kema GHJ. 2007. Electrophoretic and cytological karyotyping of the foliar wheat pathogen Mycosphaerella graminicola reveals many chromosomes with a large size range. Mycologia.
Meiklejohn CD, Montooth KL, Rand DM. 2007. Positive and negative selection on the mitochondrial genome. Trends Genet.
Menardo F, Praz CR, Wyder S, Bourras S. A, McNally KEKE, Parlange F, Riba A, Roffler S, Schaefer LK, Shimizu KKK, et al. 2016. Hybridization of powdery mildew strains gives rise to pathogens on novel agricultural crop species. Nat. Genet. [Internet] 48:201–205. Available from: http://www.nature.com/doifinder/10.1038/ng.3485
Mentlak TA, Kombrink A, Shinya T, Ryder LS, Otomo I, Saitoh H, Terauchi R, Nishizawa Y, Shibuya N, Thomma BPHJ, et al. 2012. Effector-Mediated Suppression of Chitin-Triggered Immunity by Magnaporthe oryzae Is Necessary for Rice Blast Disease. Plant Cell.
Meyne J, Ratliff RL, Moyzis RK. 1989. Conservation of the human telomere sequence (TTAGGG). among vertebrates. 86:7049–7053.
Miao VP, Covert SF, Vanetten HD. 1991. A fungal gene for antibiotic resistance on a dispensable (“B”) chromosome. Science (80-.).
Milgroom MG, Jiménez-Gasco MDM, Olivares-García C, Drott MT, Jiménez-Díaz RM. 2014. Recombination between clonal lineages of the asexual fungus Verticillium dahliae detected by genotyping by sequencing. PLoS One 9.
Möller M, Habig M, Freitag M, Stukenbrock EH. 2018. Extraordinary Genome Instability and Widespread Chromosome Rearrangements During Vegetative Growth. Genetics.
Möller M, Stukenbrock EH. 2017. Evolution and genome architecture in fungal plant pathogens. Nat. Rev. Microbiol. [Internet] 15:756–771. Available from: http://dx.doi.org/10.1038/nrmicro.2017.76
Moonen AC, Bàrberi P. 2008. Functional biodiversity: An agroecosystem approach. Agric. Ecosyst. Environ.
Muller HJ. 1932. Some Genetic Aspects of Sex. Am. Nat. [Internet] 66:118–138. Available from: http://www.journals.uchicago.edu/doi/10.1086/280418
Muller HJ. 1964. The relation of recombination to mutational advance. Mutat. Res. - Fundam. Mol. Mech. Mutagen.
Muse S V, Gaut BS. 1994. A likelihood approach for comparing synonymous and nonsynonymous nucleotide substitution rates, with application to the chloroplast genome. Mol. Biol. Evol. [Internet] 11:715–724. Available from: http://mbe.oxfordjournals.org/content/11/5/715.abstract
MYERS EW. 1995. Toward Simplifying and Accurately Formulating Fragment Assembly. J. Comput. Biol. [Internet] 2:275–290. Available from: http://www.liebertonline.com/doi/abs/10.1089/cmb.1995.2.275
Nei M, Suzuki Y, Nozawa M. 2010. The Neutral Theory of Molecular Evolution in the Genomic Era. Annu. Rev. Genomics Hum. Genet. [Internet] 11:265–289. Available from: https://doi.org/10.1146/annurev-genom-082908-150129

Nielsen R, Yang Z. 1998. Likelihood models for detecting positively selected amino acid sites and applications to the HIV-1 envelope gene. Genetics.
Nieuwenhuis BPS, James TY. 2016. The frequency of sex in fungi. Philos. Trans. R. Soc. B Biol. Sci. 371.
Notredame C, Higgins DG, Heringa J. 2000. T-Coffee: A novel method for fast and accurate multiple sequence alignment. J. Mol. Biol. [Internet] 302:205–217. Available from: http://www.sciencedirect.com/science/article/pii/S0022283600940427
Oerke E-CC. 2006. Crop losses to pests. J. Agric. Sci. 144:31–43.
Ohm R, Goodwin S, Grigoriev I, Consortium D. 2013. Diverse lifestyles and strategies of plant pathogenesis encoded in the genomes of eighteen Dothideomycetes fungi. Phytopathology 103:106.
Ohm RA, Feau N, Henrissat B, Schoch CL, Horwitz BA, Barry KW, Condon BJ, Copeland AC, Dhillon B, Glaser F, et al. 2012. Diverse Lifestyles and Strategies of Plant Pathogenesis Encoded in the Genomes of Eighteen Dothideomycetes Fungi. PLoS Pathog. 8.
Okmen B, Etalo DW, Joosten MHAJ, Bouwmeester HJ, de Vos RCH, Collemare J, De Wit PJGM. 2013. Detoxification of α-tomatine by Cladosporium fulvum is required for full virulence on tomato. New Phytol.
Okonechnikov K, Conesa A, García-Alcalde F. 2015. Qualimap 2: Advanced multi-sample quality control for high-throughput sequencing data. Bioinformatics 32:292–294.
Okonechnikov K, Golosova O, Fursov M, Varlamov A, Vaskin Y, Efremov I, German Grehov OG, Kandrov D, Rasputin K, Syabro M, et al. 2012. Unipro UGENE: A unified bioinformatics toolkit. Bioinformatics.
Ota T, Kimura M. 1971. Linkage disequilibrium between two segregating nucleotide sites under the steady flux of mutations in a finite population. Genetics.
Otto SP, Barton NH. 1997. The Evolution of Recombination: Removing the Limits to Natural Selection. Genetics [Internet] 147:879–906. Available from: http://www.genetics.org/content/147/2/879
Paoletti M, Rydholm C, Schwier EU, Anderson MJ, Szakacs G, Lutzoni F, Debeaupuis JP, Latgé JP, Denning DW, Dyer PS. 2005. Evidence for sexuality in the opportunistic fungal pathogen Aspergillus fumigatus. Curr. Biol.
Parsch J, Meiklejohn CD, Haitl DL. 2001. Patterns of DNA sequence variation suggest the recent action of positive selection in the janus-ocnus region of Drosophila simulans. Genetics 159:647–657.
Peltonen L, Pekkarinen P, Aaltonen J. 1995. Messages from an isolate: lessons from the Finnish gene pool. Biol. Chem. Hoppe. Seyler.
Peng Y, Leung HCM, Yiu SM, Chin FYL. 2012. IDBA-UD: A *de novo* assembler for single-cell and metagenomic sequencing data with highly uneven depth. Bioinformatics 28:1420–1428.
Petersen TN, Brunak S, Von Heijne G, Nielsen H. 2011. SignalP 4.0: Discriminating signal peptides from transmembrane regions. Nat. Methods 8:785–786.
Petit-Houdenot Y, Fudal I. 2017. Complex Interactions between Fungal Avirulence Genes and Their Corresponding Plant Resistance Genes and Consequences for Disease Resistance Management. Front. Plant Sci.
Pevzner PA, Tang H, Waterman MS. 2001. An Eulerian path approach to DNA fragment assembly. Proc. Natl. Acad. Sci. [Internet] 98:9748–9753. Available from: http://www.pnas.org/cgi/doi/10.1073/pnas.171285098
Pfeifer B, Wittelsbürger U, Ramos-Onsins SE, Lercher MJ. 2014. PopGenome: An efficient swiss army knife for population genomic analyses in R. Mol. Biol. Evol. 31:1929–1936.
Phillippy AM. 2017. New advances in sequence assembly. :0–3. Available from: http://genome.cshlp.org/content/27/5/xi.full.pdf+html
Pingali PL. 2012. Green Revolution: Impacts, limits, and the path ahead. Proc. Natl. Acad. Sci.
Plaumann PL, Schmidpeter J, Dahl M, Taher L, Koch C. 2018. A dispensable chromosome is required for virulence in the hemibiotrophic plant pathogen Colletotrichum higginsianum. Front. Microbiol.
Plissonneau C, Hartmann FE, Croll D. 2018. Pangenome analyses of the wheat pathogen *Zymoseptoria* tritici reveal the structural basis of a highly plastic eukaryotic genome. BMC Biol. 16:5.
Plissonneau C, Stürchler A, Croll D. 2016. The Evolution of Orphan Regions in Genomes of a Fungal Pathogen of Wheat. MBio 7:e01231-16.
Ponomarenko A, Goodwin SB, Kema GHJ. 2011. Septoria tritici blotch (STB) of wheat Septoria tritici blotch (STB) of wheat. Plant Heal. Instr.
Pop M. 2009. Genome assembly reborn: Recent computational challenges. Brief. Bioinform. 10:354–366.
Poppe S, Dorsheimer L, Happel P, Stukenbrock EH. 2015. Rapidly Evolving Genes Are Key Players in Host Specialization and Virulence of the Fungal Wheat Pathogen *Zymoseptoria* tritici (Mycosphaerella graminicola). PLoS Pathog. 11:1–21.
Lo Presti L, Lanver D, Schweizer G, Tanaka S, Liang L, Tollot M, Zuccaro A, Reissmann S, Kahmann

R. 2015. Fungal Effectors and Plant Susceptibility. Annu. Rev. Plant Biol. [Internet] 66:513–545. Available from: http://www.annualreviews.org/doi/10.1146/annurev-arplant-043014-114623
Prezeworski M, Coop G, Wall JD. 2005. THE SIGNATURE OF POSITIVE SELECTION ON STANDING GENETIC VARIATION. Evolution (N. Y).
Prugnolle F, De Meeûs T. 2008. The impact of clonality on parasite population genetic structure. Parasite.
Purcell S, Neale B, Todd-Brown K, Thomas L, Ferreira MAR, Bender D, Maller J, Sklar P, de Bakker PIW, Daly MJ, et al. 2007. PLINK: A Tool Set for Whole-Genome Association and Population-Based Linkage Analyses. Am. J. Hum. Genet. [Internet] 81:559–575. Available from: http://linkinghub.elsevier.com/retrieve/pii/S0002929707613524
Quaedvlieg W, Kema GHJ, Groenewald JZ, Verkley GJM, Seifbarghi S, Razavi M, Mirzadi Gohari A, Mehrabi R, Crous PW. 2011. *Zymoseptoria* gen. nov.: A new genus to accommodate Septoria-like species occurring on graminicolous hosts. Persoonia Mol. Phylogeny Evol. Fungi 26:57–69.
Quinlan AR, Clark RA, Sokolova S, Leibowitz ML, Zhang Y, Hurles ME, Mell JC, Hall IM. 2010. Genome-wide mapping and assembly of structural variant breakpoints in the mouse genome. Genome Res. 20:623–635.
Quinlan AR, Hall IM. 2010. BEDTools: A flexible suite of utilities for comparing genomic features. Bioinformatics 26:841–842.
Raffaele S, Farrer RA, Cano LM, Studholme DJ, MacLean D, Thines M, Jiang RHY, Zody MC, Kunjeti SG, Donofrio NM, et al. 2010. Genome evolution following host jumps in the irish potato famine pathogen lineage. Science (80-.).
Raffaele S, Kamoun S. 2012. Genome evolution in filamentous plant pathogens: why bigger can be better. Nat. Rev. Microbiol. [Internet] 10:417–430. Available from: http://dx.doi.org/10.1038/nrmicro2790
Rafiqi M, Ellis JG, Ludowici VA, Hardham AR, Dodds PN. 2012. Challenges and progress towards understanding the role of effectors in plant-fungal interactions. Curr. Opin. Plant Biol.
Ranwez V, Harispe S, Delsuc F, Douzery EJP. 2011. MACSE: Multiple alignment of coding SEquences accounting for frameshifts and stop codons. PLoS One 6.
Rausch T, Zichner T, Schlattl A, Stütz AM, Benes V, Korbel JO. 2012. DELLY: Structural variant discovery by integrated paired-end and split-read analysis. Bioinformatics.
Read AF. 1994. The evolution of virulence. Trends Microbiol.
Rice P, Longden L, Bleasby A. 2000. EMBOSS: The European Molecular Biology Open Software Suite. Trends Genet. 16:276–277.
Roberts RJ, Carneiro MO, Schatz MC. 2013. The advantages of SMRT sequencing. Genome Biol. [Internet] 14:405. Available from: http://genomebiology.biomedcentral.com/articles/10.1186/gb-2013-14-6-405
Rogers AR. 2014. How population growth affects linkage disequilibrium. Genetics 197:1329–1341.
Ronald J, Akey JM. 2005. Genome-wide scans for loci under selection in humans. Hum. Genomics.
Rouxel T, Grandaubert J, Hane JK, Hoede C, van de Wouw AP, Couloux A, Dominguez V, Anthouard VVV, Bally P, Bourras S, et al. 2011. Effector diversification within compartments of the Leptosphaeria maculans genome affected by Repeat-Induced Point mutations. Nat. Commun. [Internet] 2:202. Available from: http://www.nature.com/doifinder/10.1038/ncomms1189
Roze D, Barton NH. 2006. The Hill{\textendash}Robertson Effect and the Evolution of Recombination. Genetics [Internet] 173:1793–1811. Available from: http://www.genetics.org/content/173/3/1793
RStudio Team -. 2015. RStudio: Integrated Development for R. [Online] RStudio, Inc., Boston, MA URL http//www. rstudio. com.
Rudd JJ, Kanyuka K, Hassani-Pak K, Derbyshire M, Andongabo A, Devonshire J, Lysenko A, Saqi M, Desai NM, Powers SJ, et al. 2015. Transcriptome and Metabolite Profiling of the Infection Cycle of *Zymoseptoria tritici* on Wheat Reveals a Biphasic Interaction with Plant Immunity Involving Differential Pathogen Chromosomal Contributions and a Variation on the Hemibiotrophic Lifest. Plant Physiol. [Internet] 167:1158–1185. Available from: http://www.plantphysiol.org/lookup/doi/10.1104/pp.114.255927
Sabeti P, Reich D, JM H, Levine H, DJ R, SF S, Gabriel S, JV P, Patterson N, McDonald G, et al. 2002. Detecting recent positive selection in the human genome from haplotype structure. Nature 419.
Schardl CL, Craven KD. 2003. Interspecific hybridization in plant-associated fungi and oomycetes: A review. Mol. Ecol. 12:2861–2873.
Schiffels S, Durbin R. 2014. Inferring human population size and separation history from multiple genome sequences. Nat. Genet. [Internet] 46:919–925. Available from: http://dx.doi.org/10.1038/ng.3015
Schirawski J, Mannhaupt G, Münch K, Brefort T, Schipper K, Doehlemann G, Di Stasio M, Rössel N, Mendoza-Mendoza A, Pester D, et al. 2010. Pathogenicity determinants in smut fungi revealed by genome comparison. Science [Internet] 330:1546–1548. Available from:

http://science.sciencemag.org/content/330/6010/1546.abstract%5Cnhttp://www.ncbi.nlm.nih.gov/pubmed/21148393%5Cnhttp://0-www.sciencemag.org.alpha2.latrobe.edu.au/content/330/6010/1546.full
Schmidt SM, Lukasiewicz J, Farrer R, van Dam P, Bertoldo C, Rep M. 2015. Comparative genomics of *Fusarium oxysporum* f. sp. *melonis* reveals the secreted protein recognized by the Fom-2 resistance gene in melon. New Phytol. [Internet]:n/a--n/a. Available from: http://dx.doi.org/10.1111/nph.13584
Schoonbeek H-J, Raaijmakers JM, De Waard MA. 2002. Fungal ABC Transporters and Microbial Interactions in Natural Environments. Mpmi.
Schotanus K, Soyer JL, Connolly LR, Grandaubert J, Happel P, Smith KM, Freitag M, Stukenbrock EH. 2015. Histone modifications rather than the novel regional centromeres of *Zymoseptoria* tritici distinguish core and accessory chromosomes. Epigenetics Chromatin [Internet] 8:41. Available from: http://www.epigeneticsandchromatin.com/content/8/1/41
Schulze-Lefert P, Panstruga R. 2011. A molecular evolutionary concept connecting nonhost resistance, pathogen host range, and pathogen speciation. Trends Plant Sci. [Internet] 16:117–125. Available from: http://dx.doi.org/10.1016/j.tplants.2011.01.001
Seidl MF, Thomma BPHJ. 2014. Sex or no sex: Evolutionary adaptation occurs regardless. BioEssays 36:335–345.
Shaw MW. 1991. Interacting Effects of Interrupted Humid Periods and Light on Infection of Wheat Leaves by Mycosphaerella-Graminicola (Septoria-Tritici). Plant Pathol. 40:595–607.
Shi-Kunne X, Faino L, van den Berg GCM, Thomma BPHJ, Seidl MF. 2018. Evolution within the fungal genus Verticillium is characterized by chromosomal rearrangement and gene loss. Environ. Microbiol. 20:1362–1373.
Shokralla S, Spall JL, Gibson JF, Hajibabaei M. 2012. Next-generation sequencing technologies for environmental DNA research. Mol. Ecol. 21:1794–1805.
Short DPG, Gurung S, Hu X, Inderbitzin P, Subbarao K V. 2014. Maintenance of sex-related genes and the co- Occurrence of both mating types in verticillium dahliae. PLoS One 9.
Simonsen KL, Churchill GA, Aquadro CF. 1995. Properties of statistical tests of neutrality for DNA polymorphism data. Genetics.
Slatkin M. 1996. Gene genealogies within mutant allelic classes. Genetics.
Smit A, Hubley R. 2018. RepeatModeler Open-1.0. GItHub.
Smit A, Hubley R, Grenn P. 2013. RepeatMasker Open-4.0. RepeatMasker Open-4.0.5. [Internet]. Available from: http://www.repeatmasker.org
Smith NGC, Eyre-Walker A. 2002. Adaptive protein evolution in Drosophila. Nature.
Southern EM. 1975. Detection of specific sequences among DNA fragments separated by gel electrophoresis. J. Mol. Biol.
Soyer JL, Balesdent MH, Rouxel T, Dean RA. 2018. To B or not to B: a tale of unorthodox chromosomes. Curr. Opin. Microbiol.
Spanu PD, Abbott JC, Amselem J, Burgis TA, Soanes DM, Stüber K, Van Themaat EVL, Brown JKM, Butcher SA, Gurr SJ, et al. 2010. Genome expansion and gene loss in powdery mildew fungi reveal tradeoffs in extreme parasitism. Science (80-.). 330:1543–1546.
Sperschneider J, Gardiner DM, Dodds PN, Tini F, Covarelli L, Singh KB, Manners JM, Taylor JM. 2016. EffectorP: Predicting fungal effector proteins from secretomes using machine learning. New Phytol. 210:743–761.
Sperschneider J, Ying H, Dodds PN, Gardiner DM, Upadhyaya NM, Singh KB, Manners JM, Taylor JM. 2014. Diversifying selection in the wheat stem rust fungus acts predominantly on pathogen-associated gene families and reveals candidate effectors. Front. Plant Sci.
Staedler T, Haubold B, Merino C, Stephan W. 2008. … Impact of Sampling Schemes on the Site Frequency Spectrum in Non-equilibrium …. Available from: http://www.genetics.org/cgi/content/abstract/genetics.108.094904v1%5Cnpapers://74e8b4ef-c919-4470-93db-e267bd9e4a22/Paper/p8527
Stahl EA, Bishop JG. 2000. Plant-pathogen arms races at the molecular level. Curr. Opin. Plant Biol.
Stefansson TS, Mcdonald BA, Willi Y. 2013. Local adaptation and evolutionary potential along a temperature gradient in the fungal pathogen Rhynchosporium commune. Evol. Appl.
Stergiopoulos I, Gielkens MMC, Goodall SD, Venema K, De Waard MA. 2002. Molecular cloning and characterisation of three new ATP-binding cassette transporter genes from the wheat pathogen Mycosphaerella graminicola. Gene.
Stergiopoulos I, de Wit PJGM. 2009. Fungal Effector Proteins. Annu Rev Phytopathol [Internet] 47:233–263. Available from: http://www.ncbi.nlm.nih.gov/entrez/query.fcgi?cmd=Retrieve&db=PubMed&dopt=Citation&list_uids=19400631
Stergiopoulos I, Zwiers LH, De Waard MA. 2003. The ABC transporter MgAtr4 is a virulence factor of

Mycosphaerella graminicola that affects colonization of substomatal cavities in wheat leaves. Mol Plant Microbe Interact.
Stoate C, Boatman ND, Borralho RJ, Carvalho CR, De Snoo GR, Eden P. 2001. Ecological impacts of arable intensification in Europe. J. Environ. Manage.
Stukenbrock EH. 2013. Evolution, selection and isolation: A genomic view of speciation in fungal plant pathogens. New Phytol. 199:895–907.
Stukenbrock EH. 2016. The Role of Hybridization in the Evolution and Emergence of New Fungal Plant Pathogens. Phytopathology [Internet] 106:104–112. Available from: http://apsjournals.apsnet.org/doi/10.1094/PHYTO-08-15-0184-RVW
Stukenbrock EH, Banke S, Javan-Nikkhah M, McDonald BA. 2007. Origin and domestication of the fungal wheat pathogen Mycosphaerella graminicola via sympatric speciation. Mol. Biol. Evol. 24:398–411.
Stukenbrock EH, Bataillon T, Dutheil JY, Hansen TT, Li R, Zala M, Mcdonald BA, Wang J, Schierup MH, Stukenbrock EH, et al. 2011. The making of a new pathogen: Insights from comparative population genomics of the domesticated wheat pathogen Mycosphaerella graminicola and its wild sister species. Genome Res. 21:2157–2166.
Stukenbrock EH, Christiansen FB, Hansen TT, Dutheil JY, Schierup MH. 2012. Fusion of two divergent fungal individuals led to the recent emergence of a unique widespread pathogen species. Proc. Natl. Acad. Sci. U. S. A. [Internet] 109:10954–10959. Available from: http://www.pnas.org/content/109/27/10954.short
Stukenbrock EH, Croll D. 2014. The evolving fungal genome. Fungal Biol. Rev. [Internet] 28:1–12. Available from: http://dx.doi.org/10.1016/j.fbr.2014.02.001
Stukenbrock EH, Dutheil JY. 2017. Comparison of fine-scale recombination maps in fungal plant pathogens reveals dynamic recombination landscapes and intragenic hotspots. doi.org [Internet]:158907. Available from: https://www.biorxiv.org/content/early/2017/07/03/158907
Stukenbrock EH, Dutheil JY. 2018. Fine-scale recombination maps of fungal plant pathogens reveal dynamic recombination landscapes and intragenic hotspots. Genetics 208.
Stukenbrock EH, Jørgensen FG, Zala M, Hansen TT, McDonald BA, Schierup MH. 2010. Whole-genome and chromosome evolution associated with host adaptation and speciation of the wheat pathogen mycosphaerella graminicola. PLoS Genet. 6:1–13.
Stukenbrock EH, Mcdonald BA. 2007. Geographical variation and positive diversifying selection in the host-specific toxin SnToxA. Mol. Plant Pathol.
Stukenbrock EH, McDonald BA. 2008. The Origins of Plant Pathogens in Agro-Ecosystems. Annu. Rev. Phytopathol. [Internet] 46:75–100. Available from: http://www.annualreviews.org/doi/10.1146/annurev.phyto.010708.154114
Stukenbrock EH, McDonald BA. 2009. Population Genetics of Fungal and Oomycete Effectors Involved in Gene-for-Gene Interactions. Mol. Plant-Microbe Interact.
Stukenbrock EH, Quaedvlieg W, Javan-Nikhah M, Zala M, Crous PW, McDonald BA. 2012. *Zymoseptoria* ardabilia and *Z. pseudotritici*, two progenitor species of the septoria tritici leaf blotch fungus *Z. tritici* (synonym: Mycosphaerella graminicola). Mycologia.
Subramanian S. 2016. The effects of sample size on population genomic analyses - implications for the tests of neutrality. BMC Genomics.
Sved JA. 1971. Linkage disequilibrium and homozygosity of chromosome segments in finite populations. Theor. Popul. Biol.
Swift MJ, Izac AMN, Van Noordwijk M. 2004. Biodiversity and ecosystem services in agricultural landscapes - Are we asking the right questions? In: Agriculture, Ecosystems and Environment.
Tajima F. 1989. Statistical method for testing the neutral mutation hypothesis by DNA polymorphism. Genetics 123:585–595.
Talas F, McDonald BA. 2015. Genome-wide analysis of Fusarium graminearum field populations reveals hotspots of recombination. BMC Genomics [Internet] 16:996. Available from: http://www.biomedcentral.com/1471-2164/16/996
Tellier A, Moreno-Gámez S, Stephan W. 2014. Speed of adaptation and genomic footprints of host-parasite coevolution under arms race and trench warfare dynamics. Evolution (N. Y). 68:2211–2224.
Tenaillon MI, Austerlitz F, Tenaillon O. 2008. Apparent mutational hotspots and long distance linkage disequilibrium resulting from a bottleneck. J. Evol. Biol.
Tettelin H, Masignani V, Cieslewicz MJ, Donati C, Medini D, Ward NL, Angiuoli S V., Crabtree J, Jones AL, Durkin AS, et al. 2005. Genome analysis of multiple pathogenic isolates of Streptococcus agalactiae: Implications for the microbial "pan-genome." Proc. Natl. Acad. Sci. [Internet] 102:13950–13955. Available from: http://www.pnas.org/cgi/doi/10.1073/pnas.0506758102
Thrall PH. 2003. Evolution of Virulence in a Plant Host-Pathogen Metapopulation. Science (80-.). [Internet] 299:1735–1737. Available from:

http://www.sciencemag.org/cgi/doi/10.1126/science.1080070
Thrall PH, Papaïx J, Burdon JJ, Xie L, Zhan J. 2015. Playing on a Pathogen's Weakness: Using Evolution to Guide Sustainable Plant Disease Control Strategies. Annu. Rev. Phytopathol.
Thygesen K, Jørgensen LN, Jensen KS, Munk L. 2009. Spatial and temporal impact of fungicide spray strategies on fungicide sensitivity of Mycosphaerella graminicola in winter wheat. Eur. J. Plant Pathol.
Torriani SFF, Melichar JPE, Mills C, Pain N, Sierotzki H, Courbot M. 2015. *Zymoseptoria* tritici: A major threat to wheat production, integrated approaches to control. Fungal Genet. Biol.
Treangen TJ, Salzberg SL. 2012. Repetitive DNA and next-generation sequencing: Computational challenges and solutions. Nat. Rev. Genet.
Upadhyaya NM, Garnica DP, Karaoglu H, Sperschneider J, Nemri A, Xu B, Mago R, Cuomo CA, Rathjen JP, Park RF, et al. 2015. Comparative genomics of Australian isolates of the wheat stem rust pathogen Puccinia graminis f. sp. tritici reveals extensive polymorphism in candidate effector genes. Front. Plant Sci. [Internet] 5:759. Available from: https://www.frontiersin.org/article/10.3389/fpls.2014.00759
Van Valen L. 1973. A new evolutionary law. Evol. Theory 1:1–30.
Vernikos G, Medini D, Riley DR, Tettelin H. 2015. Ten years of pan-genome analyses. Curr. Opin. Microbiol. 23:148–154.
Waalwijk C, Mendes O, Verstappen ECP, De Waard MA, Kema GHJ. 2002. Isolation and characterization of the mating-type idiomorphs from the wheat septoria leaf blotch fungus Mycosphaerella graminicola. Fungal Genet. Biol.
De Waard MA. 1997. Significance of ABC transporters in fungicide sensitivity and resistance. In: Pesticide Science.
Walkowiak S, Rowland O, Rodrigue N, Subramaniam R. 2016. Whole genome sequencing and comparative genomics of closely related Fusarium Head Blight fungi: Fusarium graminearum, F. meridionale and F. asiaticum. BMC Genomics.
Wan J, Zhang XC, Stacey G. 2008. Chitin signaling and plant disease resistance. Plant Signal. Behav.
Wang Y, Coleman-Derr D, Chen G, Gu YQ. 2015. OrthoVenn: A web server for genome wide comparison and annotation of orthologous clusters across multiple species. Nucleic Acids Res.
Waples RS. 2006. A bias correction for estimates of effective population size based on linkage disequilibrium at unlinked gene loci. Conserv. Genet.
Watterson GA. 1975. On the number of segregating sites in genetical models without recombination. Theor. Popul. Biol.
Weissman DB, Hallatschek O. 2017. Minimal-assumption inference from population-genomic data. Elife.
Whitaker RJ, Grogan DW, Taylor JW. 2005. Recombination shapes the natural population structure of the hyperthermophilic archaeon Sulfolobus islandicus. Mol. Biol. Evol.
Wicker T, Oberhaensli S, Parlange F, Buchmann JP, Shatalina M, Roffler S, Ben-David R, Dolezel J, Simkova H, Schulze-Lefert P, et al. 2013. The wheat powdery mildew genome shows the unique evolution of an obligate biotroph. Nat Genet 45:1092–1096.
Williamson B, Tudzynski B, Tudzynski P, Van Kan JAL. 2007. Botrytis cinerea: The cause of grey mould disease. Mol. Plant Pathol.
Win J, Kamoun S. 2008. Adaptive evolution has targeted the C-terminal domain of the RXLR effectors of plant pathogenic oomycetes. Plant Signal. Behav.
De Wit PJGM. 2007. How plants recognize pathogens and defend themselves. Cell. Mol. Life Sci.
de Wit PJGM, van der Burgt A, Ökmen B, Stergiopoulos I, Abd-Elsalam KA, Aerts AL, Bahkali AH, Beenen HG, Chettri P, Cox MP, et al. 2012. The Genomes of the Fungal Plant Pathogens Cladosporium fulvum and Dothistroma septosporum Reveal Adaptation to Different Hosts and Lifestyles But Also Signatures of Common Ancestry. PLoS Genet. 8.
Wittenberg AHJ, van der Lee TAJ, M'Barek S Ben, Ware SB, Goodwin SB, Kilian A, Visser RGF, Kema GHJ, Schouten HJ. 2009. Meiosis drives extraordinary genome plasticity in the haploid fungal plant pathogen Mycosphaerella graminicola. PLoS One 4.
Wollenberg E, Vermeulen SJ, Girvetz E, Loboguerrero AM, Ramirez-Villegas J. 2016. Reducing risks to food security from climate change. Glob. Food Sec.
Wolpert TJ, Dunkle LD, Ciuffetti LM. 2002. HOST-SELECTIVE TOXINS AND AVIRULENCE DETERMINANTS: What's in a Name? *. Annu. Rev. Phytopathol 40:251–285.
Wootton JC, Feng X, Ferdig MT, Cooper RA, Mu J, Baruch DI, Magill AJ, Su XZ. 2002. Genetic diversity and chloroquine selective sweeps in Plasmodium falciparum. Nature 418:320–323.
Van De Wouw M, Kik C, Van Hintum T, Van Treuren R, Visser B. 2010. Genetic erosion in crops: Concept, research results and challenges. Plant Genet. Resour. Characterisation Util.
Wyk S Van, Wingfield BD, Vos L De, Santana QC, Merwe NA Van Der, Emma T. 2018. Multiple independent origins for a subtelomeric locus associated with growth rate in Fusarium circinatum. 9:27–36.

Yang J, Roy A, Zhang Y. 2013a. Protein-ligand binding site recognition using complementary binding-specific substructure comparison and sequence profile alignment. Bioinformatics.
Yang J, Roy A, Zhang Y. 2013b. BioLiP: A semi-manually curated database for biologically relevant ligand-protein interactions. Nucleic Acids Res.
Yang Z. 1997. PAML: a program package for phylogenetic analysis by maximum likelihood. Comput. Appl. Biosci. [Internet] 13:555–556. Available from: http://www.ncbi.nlm.nih.gov/pubmed/9367129
Yang Z. 1998. Likelihood ratio tests for detecting positive selection and application to primate lysozyme evolution. Mol. Biol. Evol.
Yang Z. 2007. PAML 4: Phylogenetic analysis by maximum likelihood. Mol. Biol. Evol. 24:1586–1591.
Yang Z, Bielawski JR. 2000. Statistical methods for detecting molecular adaptation. Trends Ecol. Evol.
Yang Z, Nielsen R. 1998. Synonymous and nonsynonymous rate variation in nuclear genes of mammals. J. Mol. Evol.
Yang Z, Nielsen R. 2002. Codon-substitution models for detecting molecular adaptation at individual sites along specific lineages. Mol. Biol. Evol.
Yang Z, Nielsen R, Goldman N, Pedersen AMK. 2000. Codon-substitution models for heterogeneous selection pressure at amino acid sites. Genetics.
Yang Z, Wong WSW, Nielsen R. 2005. Bayes empirical Bayes inference of amino acid sites under positive selection. Mol. Biol. Evol.
Yoshino K, Irieda H, Sugimoto F, Yoshioka H, Okuno T, Takano Y. 2012. Cell Death of Nicotiana benthamiana Is Induced by Secreted Protein NIS1 of Colletotrichum orbiculare and Is Suppressed by a Homologue of CgDN3. 25:625–636.
Zaffarano PL, McDonald BA, Linde CC. 2008. Rapid speciation following recent host shifts in the plant pathogenic fungus Rhynchosporium. Evolution (N. Y).
Zeder MA. 2011. The Origins of Agriculture in the Near East. Curr. Anthropol. [Internet] 52:S221–S235. Available from: https://doi.org/10.1086/659307
Zhan J, McDonald BA. 2011. Thermal adaptation in the fungal pathogen Mycosphaerella graminicola. Mol. Ecol.
Zhan J, Mundt CC, Hoffer ME, McDonald BA. 2002. Local adaptation and effect of host genotype on the rate of pathogen evolution: An experimental test in a plant pathosystem. J. Evol. Biol. 15:634–647.
Zhan J, Pettway RE, McDonald BA. 2003. The global genetic structure of the wheat pathogen Mycosphaerella graminicola is characterized by high nuclear diversity, low mitochondrial diversity, regular recombination, and gene flow. Fungal Genet. Biol.
Zheng Y, Zhang G, Lin F, Wang Z, Jin G, Yang L, Wang Y, Chen X, Xu Z, Zhao X, et al. 2008. Development of microsatellite markers and construction of genetic map in rice blast pathogen Magnaporthe grisea. Fungal Genet. Biol. 45:1340–1347.
Zhong Z, Marcel TC, Hartmann FE, Ma X, Plissonneau C, Zala M, Ducasse A, Confais J, Compain J, Lapalu N. 2017. A small secreted protein in *Zymoseptoria* tritici is responsible for avirulence on wheat cultivars carrying the Stb6 resistance gene. New Phytol.
Zipfel C, Felix G. 2005. Plants and animals: A different taste for microbes? Curr. Opin. Plant Biol. 8:353–360.
Zolan ME, Lupetti A, Guzzi G, Paladini A, Swart K, Campa M, Herschleb J, Ananiev G, Schwartz DC. 1995. Chromosome-length polymorphism in fungi. Microbiol. Rev.
Zöllner S. 2008. Population History and Linkage Disequilibrium. In: eLS. American Cancer Society. Available from: https://onlinelibrary.wiley.com/doi/abs/10.1002/9780470015902.a0005428.pub2
Zwiers LH, Stergiopoulos I, Gielkens MMC, Goodall SD, De Waard MA. 2003. ABC transporters of the wheat pathogen Mycosphaerella graminicola function as protectants against biotic and xenobiotic toxic compounds. Mol. Genet. Genomics.
Zwiers LH, De Waard MA. 2000. Characterization of the ABC transporter genes MgAtr1 and MgAtr2 from the wheat pathogen Mycosphaerella graminicola. Fungal Genet. Biol.

8. Supplement

8.1. Media and buffers

0.5M EDTA (pH 8.0):

186.1	g	EDTA•2H_2O
800	mL	H_2O

Adjust to pH 8.0 by adding NaOH pellets.

PFGE lysis buffer:

1	%	SDS
0.45	M	EDTA
1.5	mg/mL	Proteinase K (Roth)

TAE buffer (50x):

50	mM	EDTA Sodium Salt
2	M	Tris
1	M	Acetic acid

TE buffer:

89	mM	Tris base
2	mM	EDTA
80	mM	Boric acid

Tris-HCl (pH 8.0):

1	M	Tris base
800	mL	H_2O

Adjust the pH to 8.0 by adding concentrated HCl.

Yeast malt sucrose medium (YMS) - liquid:

0.4% [w/v] yeast extract

0.4% [w/v] malt extract

0.4% [w/v] sucrose

in H_2O

8.2. *Zymoseptoria* isolates

Table S 1 - This supplementary table provides detailed information about the collection year, the original ID, the sampling location, and the host species of the *Zymoseptoria* used in this study.

Species	ID	Original ID	Collection	Origin	Host	Reference
Z. ardabiliae	Za17	STIR04_1.1.1	2004	Iran	*Lolium* sp.	Stukenbrock et al. 2011
Z. ardabiliae	Za18	STIR04_1.1.2	2004	Iran	*Lolium perenne*	Stukenbrock, Quaedvlieg, et al. 2012
Z. ardabiliae	Za19	STIR04_3.3.2	2004	Iran	*Agropyron* sp.	Stukenbrock, Quaedvlieg, et al. 2012
Z. ardabiliae	Za20	STIR04_3.13.1	2004	Iran	*Agropyron* sp.	Stukenbrock et al. 2011
Z. ardabiliae	Za90	STIR11_1.1.1	2011	Iran	*Agropyron tauri*	Stukenbrock, Christiansen, et al. 2012
Z. ardabiliae	Za91	STIR11_11.4.1	2011	Iran	*Agropyron tauri*	Stukenbrock, Christiansen, et al. 2012
Z. ardabiliae	Za92	STIR11_1.10.1	2011	Iran	*Agropyron tauri*	Stukenbrock, Christiansen, et al. 2012
Z. ardabiliae	Za93	STIR11_1.15.2	2011	Iran	*Agropyron tauri*	Stukenbrock, Christiansen, et al. 2012
Z. ardabiliae	Za94	STIR11_1.15.1	2011	Iran	*Agropyron tauri*	Stukenbrock, Christiansen, et al. 2012
Z. ardabiliae	Za95	STIR11_6.1.1	2011	Iran	*Agropyron tauri*	Stukenbrock, Christiansen, et al. 2012
Z. ardabiliae	Za96	STIR11_6.1.2	2011	Iran	*Agropyron tauri*	Stukenbrock, Christiansen, et al. 2012
Z. ardabiliae	Za97	STIR11_7.5.2	2011	Iran	*Dactylis glomerata*	Stukenbrock, Christiansen, et al. 2012
Z. ardabiliae	Za98	STIR11_7.2.1	2011	Iran	*Dactylis glomerata*	Stukenbrock, Christiansen, et al. 2012
Z. ardabiliae	Za99	STIR11_7.2.4	2011	Iran	*Dactylis glomerata*	Stukenbrock, Christiansen, et al. 2012
Z. ardabiliae	Za100	STIR11_8.1.1	2011	Iran	*Agropyron tauri*	Stukenbrock, Christiansen, et al. 2012
Z. ardabiliae	Za101	STIR11_8.5.3	2011	Iran	*Agropyron tauri*	Stukenbrock, Christiansen, et al. 2012
Z. ardabiliae	Za102	STIR11_11.1.1	2011	Iran	*Agropyron tauri*	Stukenbrock, Christiansen, et al. 2012
Z. brevis	Zb85	CPC 18102	2009	Iran	*Phalaris paradoxa*	Quaedvlieg et al. 2011
Z. brevis	Zb86	CPC 18106	2009	Iran	*Phalaris minor*	Quaedvlieg et al. 2011
Z. brevis	Zb87	CPC 18110	2009	Iran	*Phalaris paradoxa*	Quaedvlieg et al. 2011
Z. brevis	Zb88	CPC 18114	2009	Iran	*Phalaris paradoxa*	Quaedvlieg et al. 2011
Z. brevis	Zb159	CPC 18107	2009	Iran	*Phalaris minor*	Quaedvlieg et al. 2011
Z. brevis	Zb161	CPC 18111	2009	Iran	*Phalaris paradoxa*	Quaedvlieg et al. 2011
Z. brevis	Zb162	CPC 18112	2009	Iran	*Phalaris paradoxa*	Quaedvlieg et al. 2011
Z. brevis	Zb163	CPC 18113	2009	Iran	*Phalaris paradoxa*	Quaedvlieg et al. 2011
Z. brevis	Zb164	CPC 18115	2009	Iran	*Phalaris paradoxa*	Quaedvlieg et al. 2011
Z. passerinii	Zpa63	S_passerinii_P63	1995	USA	*Horedum vulgare*	Stukenbrock et al. 2007
Z. pseudotritici	Zp12	ST04IR_3.11.1	2004	Iran	*Elymus repens*	Stukenbrock et al. 2011
Z. pseudotritici	Zp13	ST04IR_2.2.1	2004	Iran	*Dactylis glomerata*	Stukenbrock et al. 2011
Z. pseudotritici	Zp14	ST04IR_4.3.1	2004	Iran	*Elymus repens*	Stukenbrock et al. 2011
Z. pseudotritici	Zp15	ST04IR_5.3	2004	Iran	*Dactylis glomerata*	Stukenbrock et al. 2011
Z. pseudotritici	Zp16	ST04IR_5.9.1	2004	Iran	*Dactylis glomerata*	Stukenbrock et al. 2011
Z. pseudotritici	Zp36	ST04IR_2.2.3	2004	Iran	*Dactylis glomerata*	Stukenbrock, Christiansen, et al. 2012
Z. pseudotritici	Zp37	ST04IR_2.2.4	2004	Iran	*Dactylis glomerata*	Stukenbrock, Christiansen, et al. 2012
Z. pseudotritici	Zp60	ST04IR_3.12.1	2004	Iran	*Elymus repens*	Stukenbrock, Christiansen, et al. 2012
Z. pseudotritici	Zp65	ST04IR_4.1.1	2004	Iran	*Elymus repens*	Stukenbrock, Christiansen, et al. 2012
Z. pseudotritici	Zp66	ST04IR_4a.1.2	2004	Iran	*Elymus repens*	Stukenbrock, Christiansen, et al. 2012
Z. pseudotritici	Zp67	ST04IR_4b.1.2	2004	Iran	*Elymus repens*	Stukenbrock, Christiansen, et al. 2012
Z. pseudotritici	Zp68	ST04IR_4b.1.3	2004	Iran	*Elymus repens*	Stukenbrock, Christiansen, et al. 2012
Z. pseudotritici	Zp69	ST04IR_4b.1.4	2004	Iran	*Elymus repens*	Stukenbrock, Christiansen, et al. 2012
Z. pseudotritici	Zp70	ST04IR_4.3.2	2004	Iran	*Elymus repens*	Stukenbrock, Christiansen, et al. 2012
Z. pseudotritici	Zp71	ST04IR_4.7.2	2004	Iran	*Elymus repens*	Stukenbrock, Christiansen, et al. 2012
Z. pseudotritici	Zp72	ST04IR_4.8.1	2004	Iran	*Elymus repens*	Stukenbrock, Christiansen, et al. 2012
Z. pseudotritici	Zp73	ST04IR_5.2	2004	Iran	*Dactylis glomerata*	Stukenbrock, Christiansen, et al. 2012
Z. pseudotritici	Zp74	ST04IR_5.5	2004	Iran	*Dactylis glomerata*	Stukenbrock, Christiansen, et al. 2012
Z. pseudotritici	Zp75	ST04IR_5.6	2004	Iran	*Dactylis glomerata*	Stukenbrock, Christiansen, et al. 2012
Z. pseudotritici	Zp76	ST04IR_5.7.1	2004	Iran	*Dactylis glomerata*	Stukenbrock, Christiansen, et al. 2012
Z. pseudotritici	Zp77	ST04IR_5.7.2	2004	Iran	*Dactylis glomerata*	Stukenbrock, Christiansen, et al. 2012
Z. pseudotritici	Zp78	ST04IR_5.8.1	2004	Iran	*Dactylis glomerata*	Stukenbrock, Christiansen, et al. 2012
Z. pseudotritici	Zp79	ST04IR_5.8.2	2004	Iran	*Dactylis glomerata*	Stukenbrock, Christiansen, et al. 2012
Z. pseudotritici	Zp80	ST04IR_5.8	2004	Iran	*Dactylis glomerata*	Stukenbrock, Christiansen, et al. 2012
Z. pseudotritici	Zp81	ST04IR_5.9.2	2004	Iran	*Dactylis glomerata*	Stukenbrock, Christiansen, et al. 2012
Z. pseudotritici	Zp82	ST04IR_5.10.1	2004	Iran	*Dactylis glomerata*	Stukenbrock, Christiansen, et al. 2012
Z. pseudotritici	Zp83	ST04IR_5.10.2	2004	Iran	*Dactylis glomerata*	Stukenbrock, Christiansen, et al. 2012
Z. tritici	Zt02	MgDk09_F24	2009	Denmark	*Triticum aestivum*	Thygesen et al. 2009
Z. tritici	Zt04	MgDK09_U33	2009	Denmark	*Triticum aestivum*	Thygesen et al. 2009
Z. tritici	Zt05	MgDk09_U34	2009	Denmark	*Triticum aestivum*	Thygesen et al. 2009
Z. tritici	Zt07	MgDk09_U43	2009	Denmark	*Triticum aestivum*	Thygesen et al. 2009
Z. tritici	Zt09	IPO323	1995	Holland	*Triticum aestivum*	Kema & Silfhout 1997
Z. tritici	Zt10	ST04IR_A26b	2004	Iran	*Triticum aestivum*	Stukenbrock et al. 2007
Z. tritici	Zt11	ST04IR_48b	2004	Iran	*Triticum aestivum*	Stukenbrock et al. 2007
Z. tritici	Zt148	Zt_Ger_13_1_1_2	2013	Germany	*Triticum aestivum*	Grandaubert et al. 2017
Z. tritici	Zt150	Zt_Ger_13_2_1_1	2013	Germany	*Triticum aestivum*	Grandaubert et al. 2017
Z. tritici	Zt151	Zt_Ger_13_3_1_1	2013	Germany	*Triticum aestivum*	Grandaubert et al. 2017
Z. tritici	Zt153	Zt_Ger_13_5_1_1	2013	Germany	*Triticum aestivum*	Grandaubert et al. 2017
Z. tritici	Zt154	Zt_Ger_13_6_1_1	2013	Germany	*Triticum aestivum*	Grandaubert et al. 2017
Z. tritici	Zt155	Zt_Fr_13_1_1_1	2013	France	*Triticum aestivum*	Grandaubert et al. 2017

8.3. *De novo* genome assemblies from short-read sequencing

Table S 2 - This supplementary table provides information on the used sequencing technology and the produced *de novo* genome assemblies. The table provides assembly statistics for both the raw (light grey) and the filtered (dark grey, no contigs <1kb) versions for each fungal strain.

Species	ID	Technology	Coverage	Length (bp)	Scaf-folds	N50 (bp)	N/100kb	Length	Scaf-folds	N50	N/100kb	Assembler	Reference
Z. ardabiliae	Za18	Illumina (pe 100)	28.36 x	36,881,606	67,498	129,899	21.99	31,570,810	760	129,943	21.37	SPAdes	unpublished
Z. ardabiliae	Za19	Illumina (pe 100)	2.79 x	33,793,808	2,391	59,564	180.23	33,428,317	1,483	60,279	163.78	SPAdes	unpublished
Z. ardabiliae	Za20	Illumina (pe 100)	15.35 x	38,823,759	63,18	100,394	42.61	33,748,859	1,137	103,153	38.62	SPAdes	unpublished
Z. ardabiliae	Za90	Illumina (pe 100)	36.75 x	39,889,619	33,832	77,689	2.71	33,457,969	2,029	88,270	2.93	SPAdes	unpublished
Z. ardabiliae	Za91	Illumina (pe 100)	52.66 x	39,540,228	25,883	87,577	4.27	34,589,363	2,222	94,224	4.53	SPAdes	unpublished
Z. ardabiliae	Za92	Illumina (pe 100)	49.99 x	41,102,514	31,183	76,464	7.67	34,955,877	2,61	82,534	8.25	SPAdes	unpublished
Z. ardabiliae	Za93	Illumina (pe 100)	35.01 x	41,516,589	38,793	74,028	5.87	34,445,610	2,069	80,211	6.31	SPAdes	unpublished
Z. ardabiliae	Za94	Illumina (pe 100)	32.46 x	41,899,077	40,651	76,321	6.12	34,250,062	1,941	84,520	6.6	SPAdes	unpublished
Z. ardabiliae	Za95	Illumina (pe 100)	52.08 x	38,548,600	30,229	109,127	2.45	33,372,619	1,43	118,933	2.58	SPAdes	unpublished
Z. ardabiliae	Za96	Illumina (pe 100)	36.05 x	38,533,166	33,025	99,35	1.37	33,069,931	1,226	105,534	1.37	SPAdes	unpublished
Z. ardabiliae	Za97	Illumina (pe 100)	51.72 x	39,467,009	26,151	101,521	6.97	34,377,111	2,061	111,537	7.38	SPAdes	unpublished
Z. ardabiliae	Za98	Illumina (pe 100)	31.82 x	39,613,409	32,643	92,398	5.56	33,712,061	1,649	97,015	5.91	SPAdes	unpublished
Z. ardabiliae	Za99	Illumina (pe 100)	35.56 x	38,991,847	28,638	92,458	8.21	33,541,388	1,726	98,766	8.21	SPAdes	unpublished
Z. ardabiliae	Za100	Illumina (pe 100)	51.84 x	39,674,999	23,441	97,225	5.1	35,253,837	1,865	101,581	5.35	SPAdes	unpublished
Z. ardabiliae	Za101	Illumina (pe 100)	33.06 x	39,450,626	25,977	84,17	5.33	34,866,733	1,732	90,151	5.58	SPAdes	unpublished
Z. ardabiliae	Za102	Illumina (pe 100)	46.22 x	38,767,929	26,992	95,563	2.47	33,818,640	1,898	103,384	2.61	SPAdes	unpublished
Z. brevis	Zb85	Illumina (pe 100)	49.30 x	33,890,224	14,191	149,377	597.47	31,663,403	865	150,007	405.82	SOAPdenovo 2	unpublished
Z. brevis	Zb86	Illumina (pe 100)	35.73 x	33,458,074	10,249	78,818	1,508.55	31,330,229	1,276	82,090	972.57	SOAPdenovo 2	unpublished
Z. brevis	Zb88	Illumina (pe 100)	69.72 x	35,279,156	12,339	69,466	2,816.29	32,493,593	1,845	71,721	1,961.53	SOAPdenovo 2	unpublished

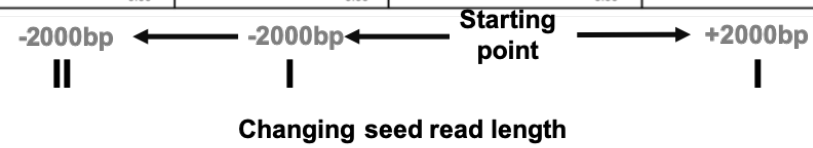
Z. brevis	Zb159	Illumina (pe 100)	54.04 x	35,669,403	30,505	92,84	1,165.98	30,673,595	1,446	96,153	744.42	SOAPdenovo 2	unpublished
Z. brevis	Zb161	Illumina (pe 100)	54.45 x	37,645,874	30,779	63,08	2,216.71	32,680,433	2,033	66,360	1,526.55	SOAPdenovo 2	unpublished
Z. brevis	Zb162	Illumina (pe 100)	61.71 x	39,521,122	36,687	74,355	840.64	33,374,479	2,038	79,511	603.31	SOAPdenovo 2	unpublished
Z. brevis	Zb163	Illumina (pe 100)	58.03 x	39,765,478	37,289	69,481	1,013.94	33,872,881	1,997	72,605	701.82	SOAPdenovo 2	Grandaubert et al. 2015
Z. brevis	Zb164	Illumina (pe 100)	56.71 x	35,531,974	15,345	66,118	980.77	32,390,191	1,659	68,501	628.26	SOAPdenovo 2	unpublished
Z. pseudotritici	Zp12	Illumina (pe 100)	29.42 x	33,550,923	2,431	86,4	223.43	33,334,016	1,097	87,522	170.05	SOAPdenovo 2	unpublished
Z. pseudotritici	Zp14	Illumina (pe 100)	30.02 x	32,924,327	1,074	106,324	0	33,334,016	1,074	106,324	3581.09	SOAPdenovo 2	unpublished
Z. pseudotritici	Zp36	Illumina (pe 100)	51.93 x	34,755,683	4,237	73,781	4,272.87	33,767,936	1,274	74,096	0	SOAPdenovo 2	unpublished
Z. pseudotritici	Zp37	Illumina (pe 100)	36.48 x	34,003,222	3,753	73,806	0	33,160,207	1,222	74,080	0	SOAPdenovo 2	unpublished
Z. pseudotritici	Zp60	Illumina (pe 100)	59.07 x	34,921,205	4,101	77,73	0	33,960,069	1,242	78,036	0	SOAPdenovo 2	unpublished
Z. pseudotritici	Zp65	Illumina (pe 100)	58.90 x	34,805,853	4,155	79,866	0	33,823,543	1,199	80,348	0	SOAPdenovo 2	unpublished
Z. pseudotritici	Zp66	Illumina (pe 100)	40.55 x	34,230,561	4,26	73,704	0	33,688,589	1,225	73,777	0	SOAPdenovo 2	unpublished
Z. pseudotritici	Zp67	Illumina (pe 100)	34.35 x	34,119,667	3,75	72,071	0	33,198,762	1,258	72,376	0	SOAPdenovo 2	unpublished
Z. pseudotritici	Zp68	Illumina (pe 100)	38.94 x	34,480,703	3,928	78,622	0	33,619,621	1,205	80,346	0	SOAPdenovo 2	unpublished
Z. pseudotritici	Zp69	Illumina (pe 100)	35.82 x	34,965,056	3,811	73,208	0	34,002,180	1,266	73,783	0	SOAPdenovo 2	unpublished
Z. pseudotritici	Zp70	Illumina (pe 100)	62.59 x	34,660,754	4,16	77,398	0	33,689,613	1,29	77,897	0	SOAPdenovo 2	unpublished
Z. pseudotritici	Zp71	Illumina (pe 100)	38.07 x	34,662,937	4,146	69,027	0	33,702,102	1,259	69,707	0	SOAPdenovo 2	unpublished
Z. pseudotritici	Zp72	Illumina (pe 100)	40.18 x	35,026,958	4,149	69,05	0	33,998,147	1,268	60,707	0	SOAPdenovo 2	unpublished
Z. pseudotritici	Zp73	Illumina (pe 100)	39.78 x	34,276,931	4,315	75,365	0	33,325,077	1,276	77,662	0	SOAPdenovo 2	unpublished
Z. pseudotritici	Zp74	Illumina (pe 100)	59.00 x	34,239,919	4,07	69,342	0	33,301,895	1,231	69,577	0	SOAPdenovo 2	unpublished
Z. pseudotritici	Zp75	Illumina (pe 100)	38.14 x	34,957,142	4,077	67,28	0	33,881,675	1,280	67,635	0	SOAPdenovo 2	unpublished
Z. pseudotritici	Zp76	Illumina (pe 100)	61.91 x	34,912,172	4,422	73,601	0	33,881,675	1,306	74,002	0	SOAPdenovo 2	unpublished
Z. pseudotritici	Zp77	Illumina (pe 100)	38.00 x	34,243,643	4,365	74,934	0	33,269,950	1,324	75,531	0	SOAPdenovo 2	unpublished
Z. pseudotritici	Zp78	Illumina (pe 100)	39.19 x	33,606,148	4,042	80,652	0	32,689,316	1,19	80,839	0	SOAPdenovo 2	unpublished

Z. pseudotritici	Zp79	Illumina (pe 100)	39.43 x	34,915,389	3,948	72,51	0	33,920,353	1,217	72,888	0	SOAPdenovo 2	unpublished
Z. pseudotritici	Zp80	Illumina (pe 100)	37.39 x	34,915,389	3,948	72,51	0	33,920,353	1,217	72,888	0	SOAPdenovo 2	unpublished
Z. pseudotritici	Zp81	Illumina (pe 100)	60.29 x	34,915,389	4,134	76,306	0	33,920,353	1,232	77,237	0	SOAPdenovo 2	unpublished
Z. pseudotritici	Zp82	Illumina (pe 100)	35.41 x	39,991,689	21,855	69,035	0.01	33,884,348	1,316	71,425	0	SOAPdenovo 2	unpublished
Z. pseudotritici	Zp83	Illumina (pe 100)	38.27 x	34,848,684	4,248	71,421	0	33,851,431	1,262	72,113	0	SOAPdenovo 2	unpublished
Z. tritici	Zt02	Illumina (pe 100)	55.11 x	39,374,274	12,117	90,326	889.04	37,536,753	1,305	90,417	818.18	SOAPdenovo 2	**G**randaubert et al. 2017
Z. tritici	Zt04	Illumina (pe 100)	55.15 x	40,294,658	15,422	97,676	595.76	38,037,954	1,541	98,493	558.72	SOAPdenovo 2	**G**randaubert et al. 2017
Z. tritici	Zt07	Illumina (pe 100)	51.82 x	39,209,627	13,814	79,156	1,033.46	37,261,558	1,329	79,703	955.35	SOAPdenovo 2	**G**randaubert et al. 2017
Z. tritici	Zt11	Illumina (pe 100)	26.20 x	32,769,954	951	108,52	2964.02	32,769,954	951	108,52	2964.02	SOAPdenovo 2	**G**randaubert et al. 2017
Z. tritici	Zt148	Illumina (pe 100)	55.21 x	39,251,332	15,962	88,65	1,210.40	36,997,818	1,356	89,053	1,078.35	SOAPdenovo 2	**G**randaubert et al. 2017
Z. tritici	Zt150	Illumina (pe 100)	58.48 x	40,136,236	11,792	101,324	964.78	38,442,271	1,283	102,358	897.41	SOAPdenovo 2	**G**randaubert et al. 2017
Z. tritici	Zt151	Illumina (pe 100)	54.92 x	39,431,307	13,364	92,073	869.82	37,434,734	1,292	92,964	783.84	SOAPdenovo 2	**G**randaubert et al. 2017
Z. tritici	Zt153	Illumina (pe 100)	47.58 x	39,120,492	15,115	83,879	1,083.59	36,845,152	1,462	84,896	926.77	SOAPdenovo 2	**G**randaubert et al. 2017
Z. tritici	Zt154	Illumina (pe 100)	49.31 x	39,184,555	17,378	81,241	1,038,86	36,656,034	1,402	81,493	926.06	SOAPdenovo 2	**G**randaubert et al. 2017
Z. tritici	Zt155	Illumina (pe 100)	49.57 x	39,175,218	15,781	93,105	1,116.09	36,911,264	1,407	94,187	1,004.03	SOAPdenovo 2	**G**randaubert et al. 2017

8.4. *De novo* genome assemblies from SMRT sequencing - An example

A

	Seed read length 13 Kb	Seed read length 15 Kb	Seed read length 17 Kb	Seed read length 19 Kb
# contigs (>= 0 bp)	24	23	24	32
# contigs (>= 1000 bp)	24	23	24	32
# contigs (>= 5000 bp)	24	23	24	32
# contigs (>= 10000 bp)	24	23	24	32
# contigs (>= 25000 bp)	22	21	22	29
# contigs (>= 50000 bp)	21	20	20	22
Total length (>= 0 bp)	39445980	39425496	39458478	39687397
Total length (>= 1000 bp)	39445980	39425496	39458478	39687397
Total length (>= 5000 bp)	39445980	39425496	39458478	39687397
Total length (>= 10000 bp)	39445980	39425496	39458478	39687397
Total length (>= 25000 bp)	39413305	39386399	39426393	39616589
Total length (>= 50000 bp)	39374758	39344643	39343637	39393809
# contigs	24	23	24	32
Largest contig	6591746	6591773	6591771	6591743
Total length	39445980	39425496	39458478	39687397
GC (%)	51.69	51.70	51.67	51.66
N50	2925378	2925395	2925384	2925323
N75	2001518	2001534	1943094	1669488
L50	5	5	5	5
L75	9	9	10	10
# N's per 100 kbp	0.00	0.00	0.00	0.00



Changing seed read length

B

	Seed read length (bp)			
	13,000	15,000	17,000	19,000
Maximal number of contigs	🔴	🟢	🔴	🔴
Maximal assembly length	🔴	🔴	🔴	🟢
N50	🔴	🟢	🔴	🔴

Figure S 1 - This figure illustrates the selection process for the *de novo* genome assemblies generated with SMRTAssembly using the example of Zt10. A) This table shows the general assembly statistics of the *de novo* genome assemblies generated using the software Quast (Gurevich et al. 2013). First, the assembly procedure is done with automatic selection of the seed read length (in this case 17,000 bp). Subsequently, further assemblies are created with manually set seed read lengths at 2000bp more or less (15,000bp and 19,000bp) (I). In an improvement of the assembly result, in the next step (II), the seed read lengths were further changed according to the same scheme until the quality of the resulting assemblies decreased again (which occurred with a seed read length of 13,000bp). B) This table shows the quality rating of the different assemblies according to the characteristics "maximum number of contigs", "maximum assembly length" and "N50". The quality rating bases on the assembly with the smallest number of contigs, the shortest assembly length, and the highest N50. In this case, the assembly created with a seed read length of 15,000 shows the best result for two of the three characteristics and is therefore used as a reference genome.

8.5. Circular genome representations of Za17, Zpa63, Zp13 Zt05 and Zt05

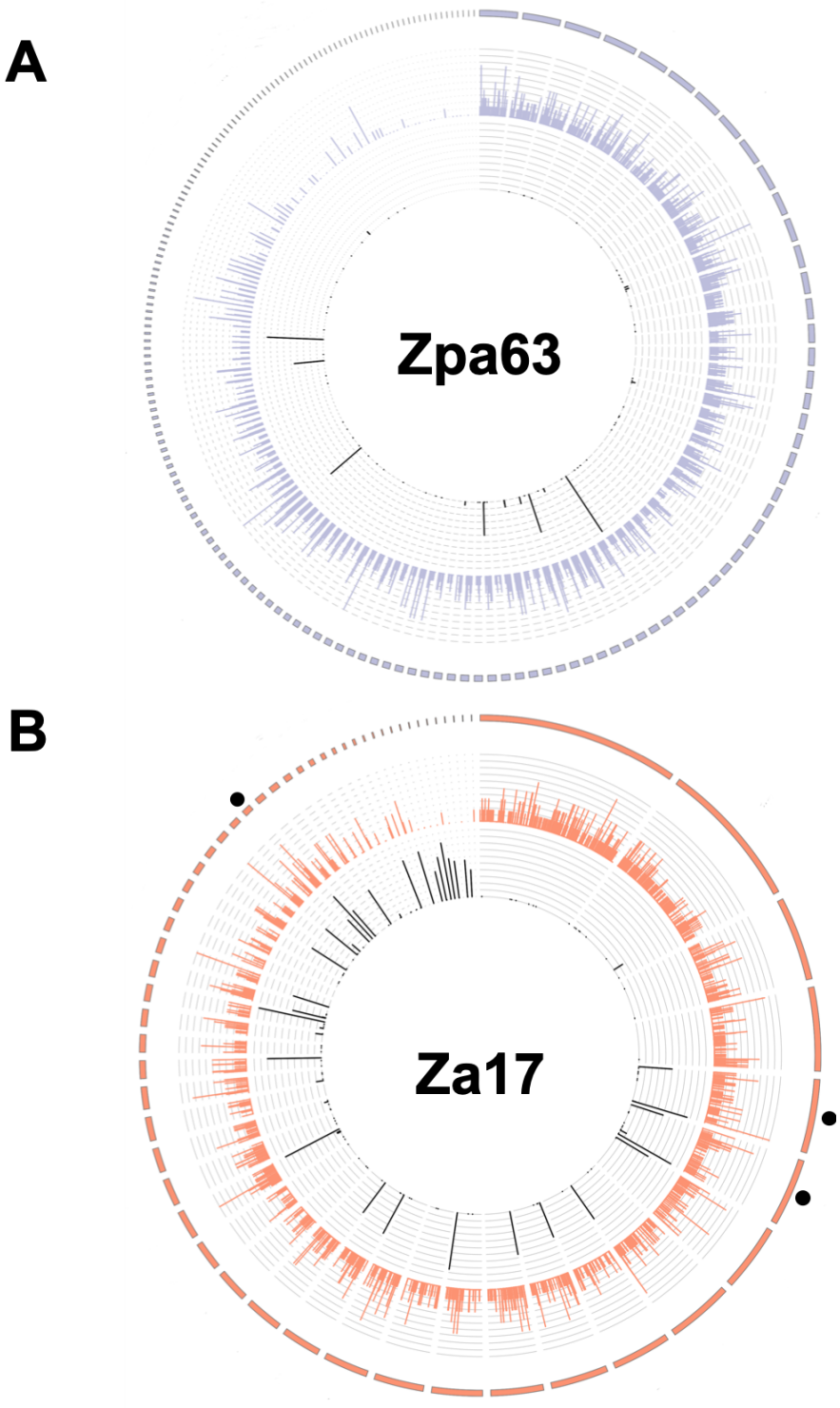


Figure S 2 - Genome representation of the *de novo* genome assemblies of reference strains Za17 (*Z. arda-biliae*) and Zpa63 (*Z. passerinii*). The black dots mark unitigs with telomeric repeats at both ends. The circular data tracks show (from inside to outside): 1. The distribution and density of the telomeric repeats, 2. The distribution and density of repetitive sequences, and 3. The length of the unitigs in decreasing order. Colored histogram track = density with repetitive sequences per window. Black histogram track = Density of telomeric repeats within a sliding window.

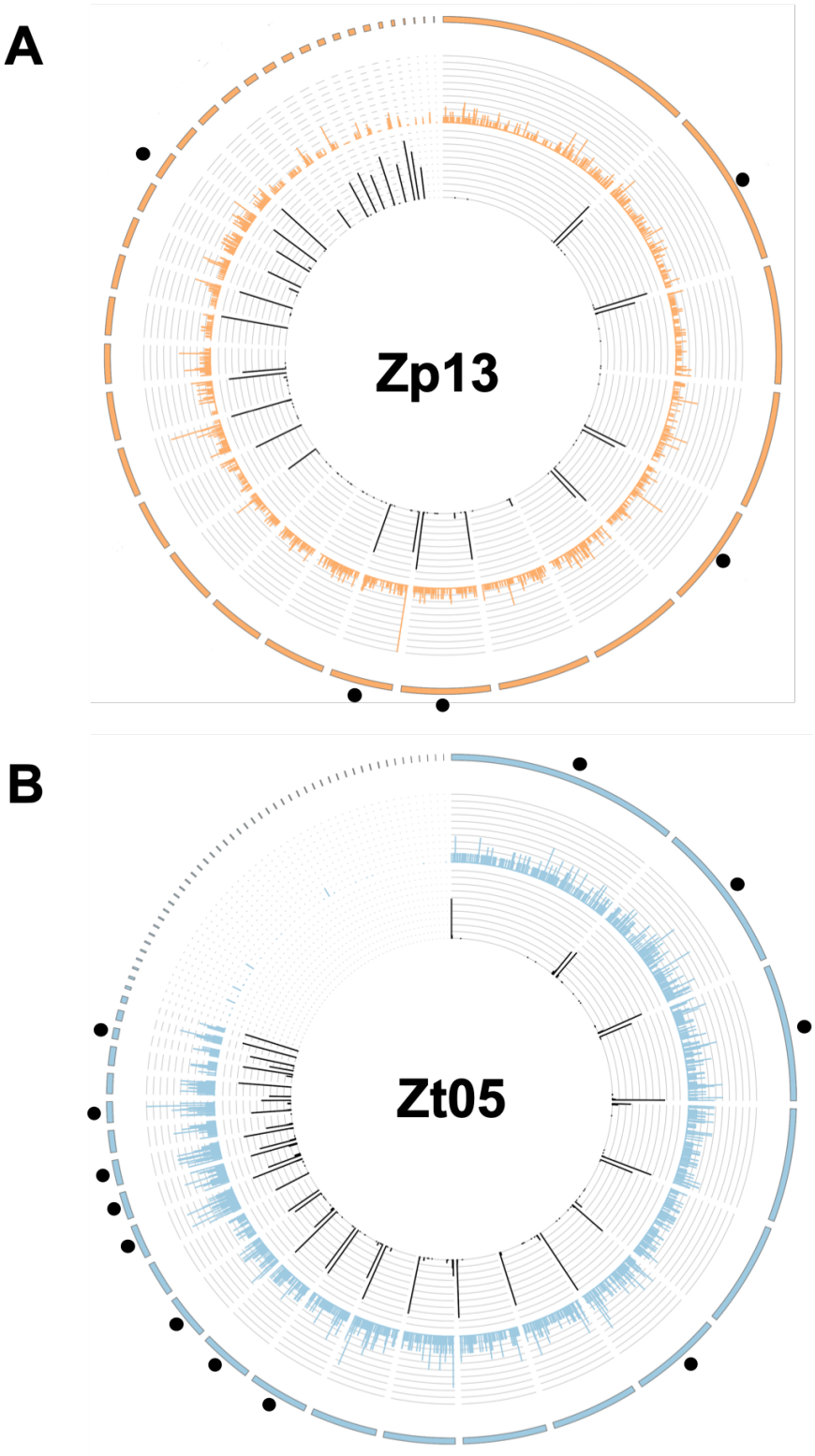


Figure S 3 - Genome representation of the *de novo* genome assemblies of reference strains Z13 (*Z. pseudotritici*) and Zt05 (*Z. tritici*). The black dots mark unitigs with telomeric repeats at both ends. The circular data tracks show (from inside to outside): 1. The distribution and density of the telomeric repeats, 2. The distribution and density of repetitive sequences, and 3. The length of the unitigs in decreasing order. Colored histogram track = density with repetitive sequences per window. Black histogram track = Density of telomeric repeats within a sliding window.

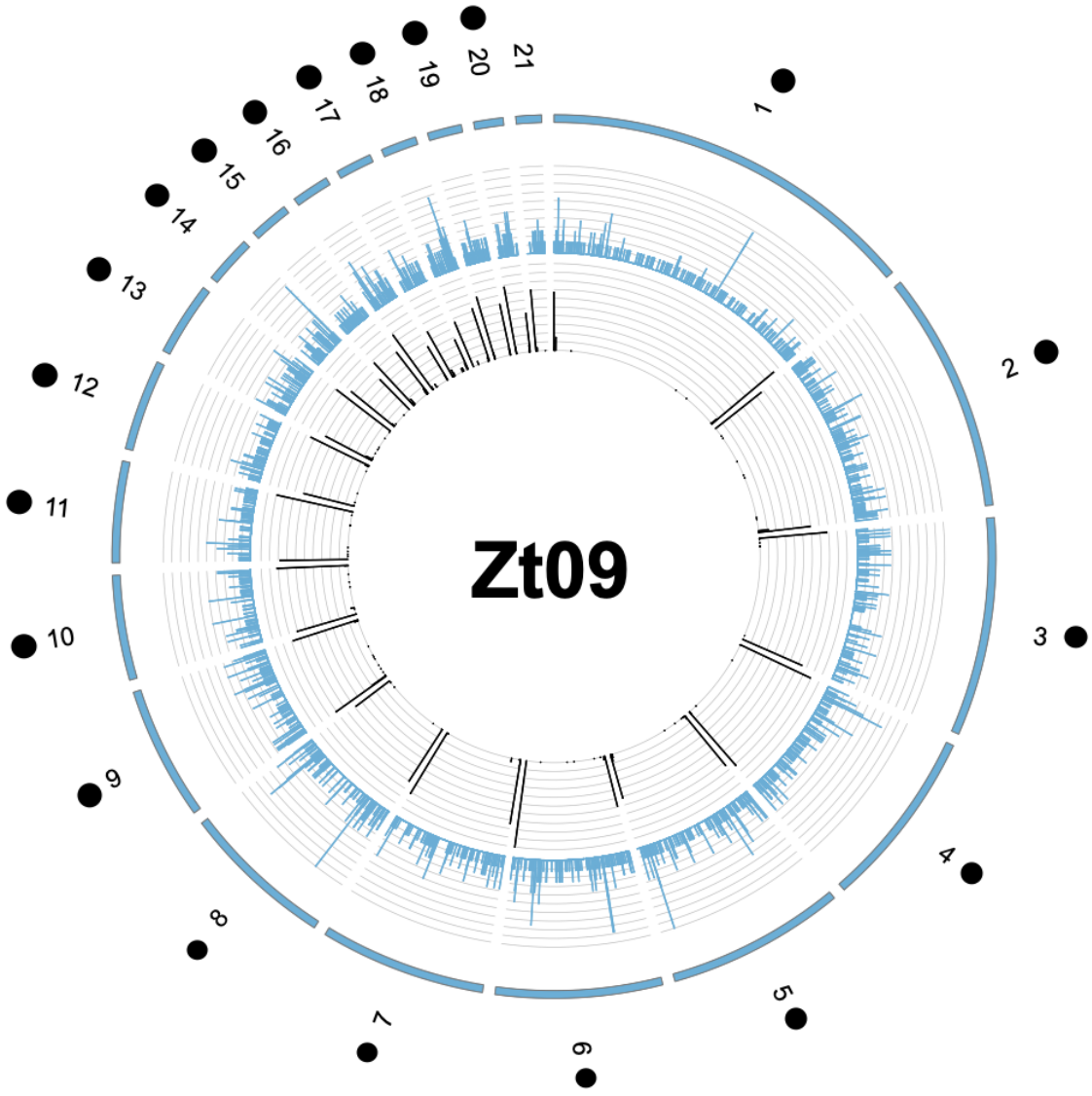


Figure S 4 - Genome representation of the *de novo* genome assemblies of reference strains Zt09 (*Z. tritici*). The black dots mark unitigs with telomeric repeats at both ends. The circular data tracks show (from inside to outside): 1. The distribution and density of the telomeric repeats, 2. The distribution and density of repetitive sequences, and 3. The length of the unitigs in decreasing order. The results confirm the finding of 20 wholly sequenced chromosomes (1-20) reported in the original publication (Goodwin et al. 2011). Colored histogram track = density with repetitive sequences per window. Black histogram track = Density of telomeric repeats within a sliding window.

8.6. Comparative analyses of synteny in the *Zymoseptoria* species complex

8.6.1. Circular and linear representation of the synteny analyses between Zt09 and Zpa63, Zp13 and Zt05

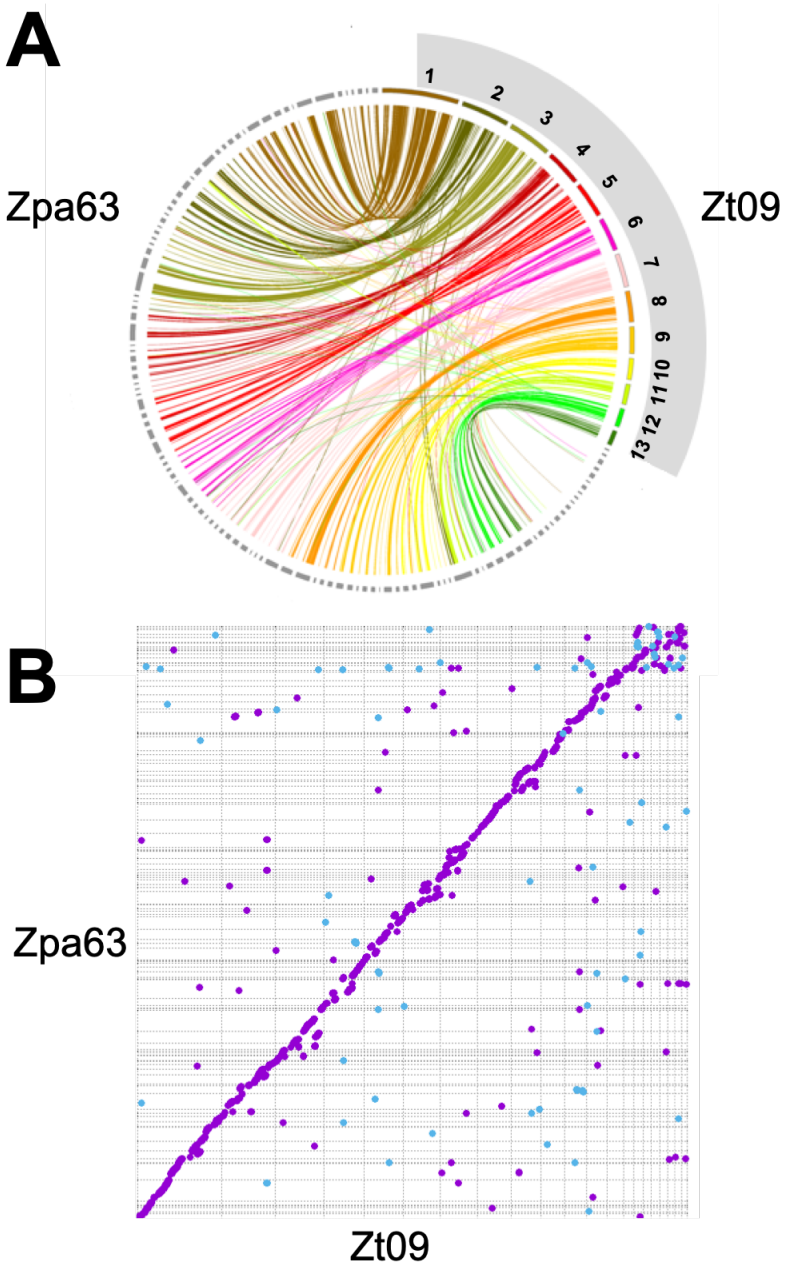


Figure S 5 - Synteny analysis between Zpa63 (*Z. passerinii*) and Zt09 (*Z. tritici*). A) The ideogram suggested a high degree of synteny of the assembly blocks from Zpa63 to Zt09. Black points indicate completely assembled unitigs (telomeric repeats at both ends) B) The synteny dot plot shows a broken diagonal indicating macrosynteny. Shorter structural variations, indicated here by deviations from the central diagonal, are also displayed.

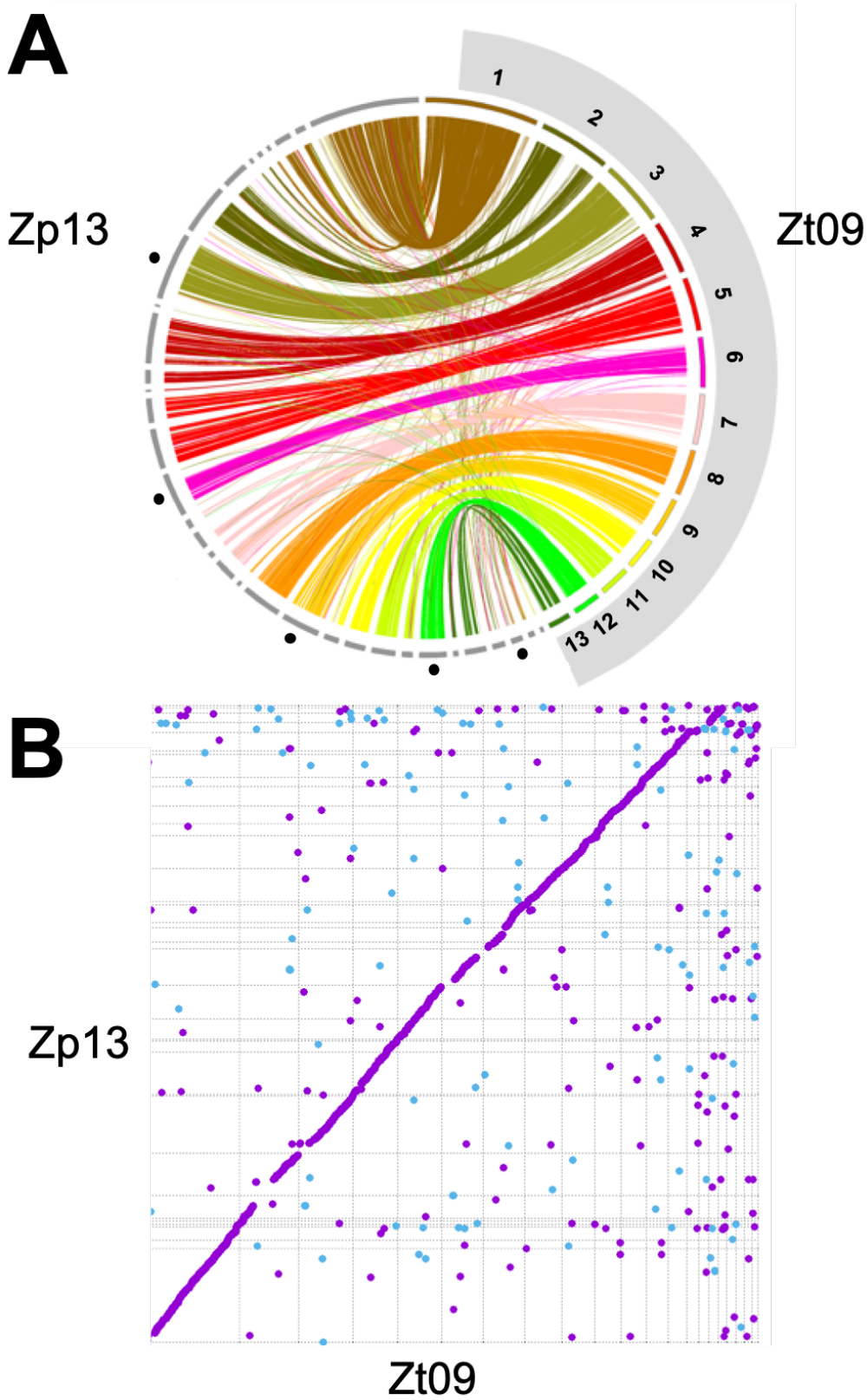


Figure S 6 - Synteny analysis between Zp13 (*Z. pseudotritici*) and Zt09 (*Z. tritici*). A) The ideogram suggested a high degree of synteny of the assembly blocks from Zpa63 to Zt09. Black points indicate completely assembled unitigs (telomeric repeats at both ends) B) The synteny dot plot shows a broken diagonal indicating macrosynteny. Shorter structural variations, indicated here by deviations from the central diagonal, are also displayed.

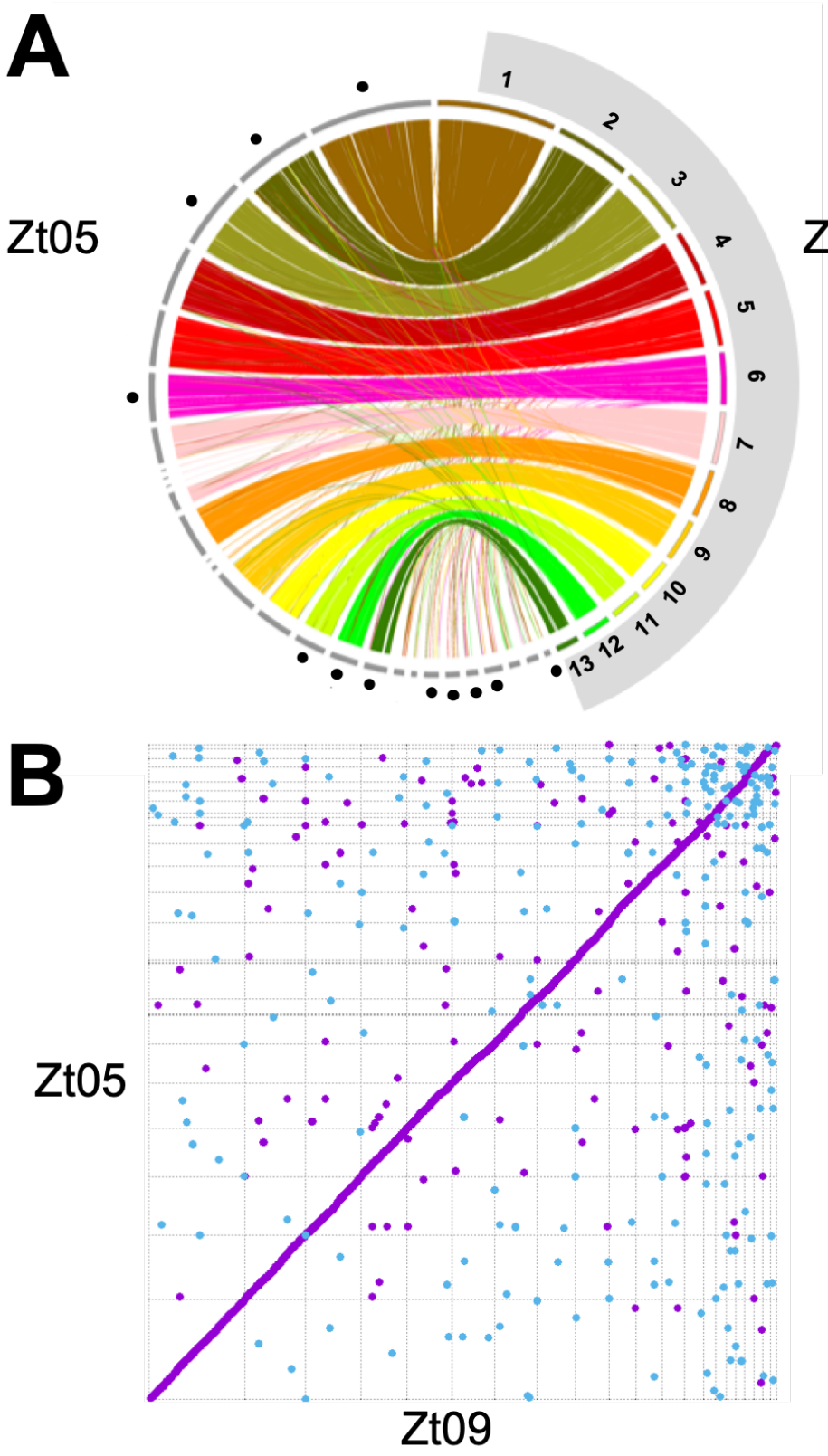


Figure S 7 - Synteny analysis between Zt05 (*Z. tritici*) and Zt09 (*Z. tritici*). A) The ideogram suggested a high degree of synteny of the assembly blocks from Zpa63 to Zt09. Black points indicate completely assembled unitigs (telomeric repeats at both ends) B) The synteny dot plot shows a broken diagonal indicating macrosynteny. Shorter structural variations, indicated here by deviations from the central diagonal, are also displayed.

8.6.2. Detailed representation of the circular synteny plots between the core genome of *Z. tritici* and the wild grass associated sister species

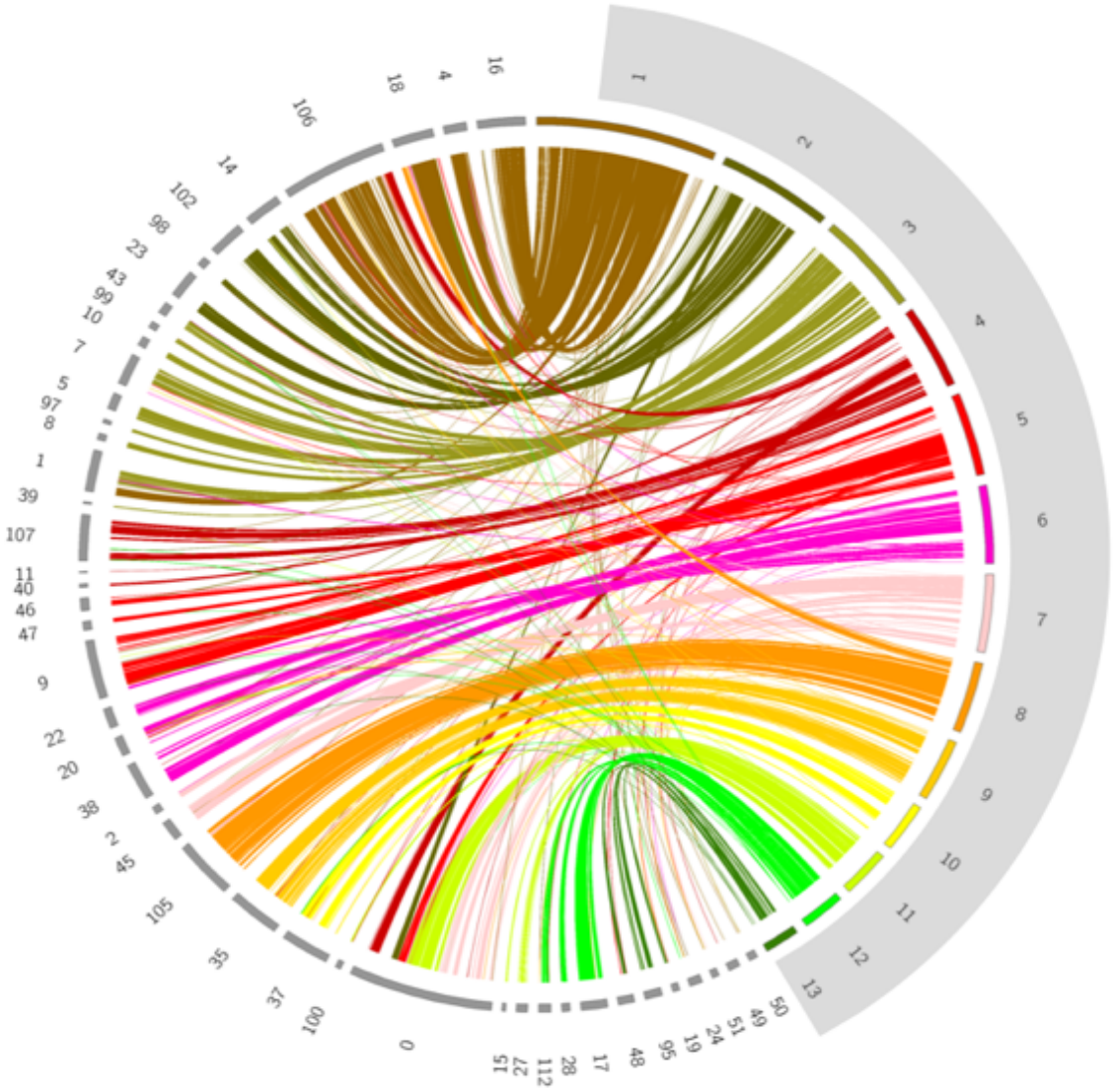


Figure S 8 - Circular synteny plot of the unitigs of Za17 and the core chromosomes of Zt09. The ideogram shows that the majority of links form large synteny blocks, but there are occasional synteny breaks due to translocations. Especially unitig0 shows many different extended synteny blocks. However, since these breaks may also be due to assembly errors, more in-depth analysis must verify them.

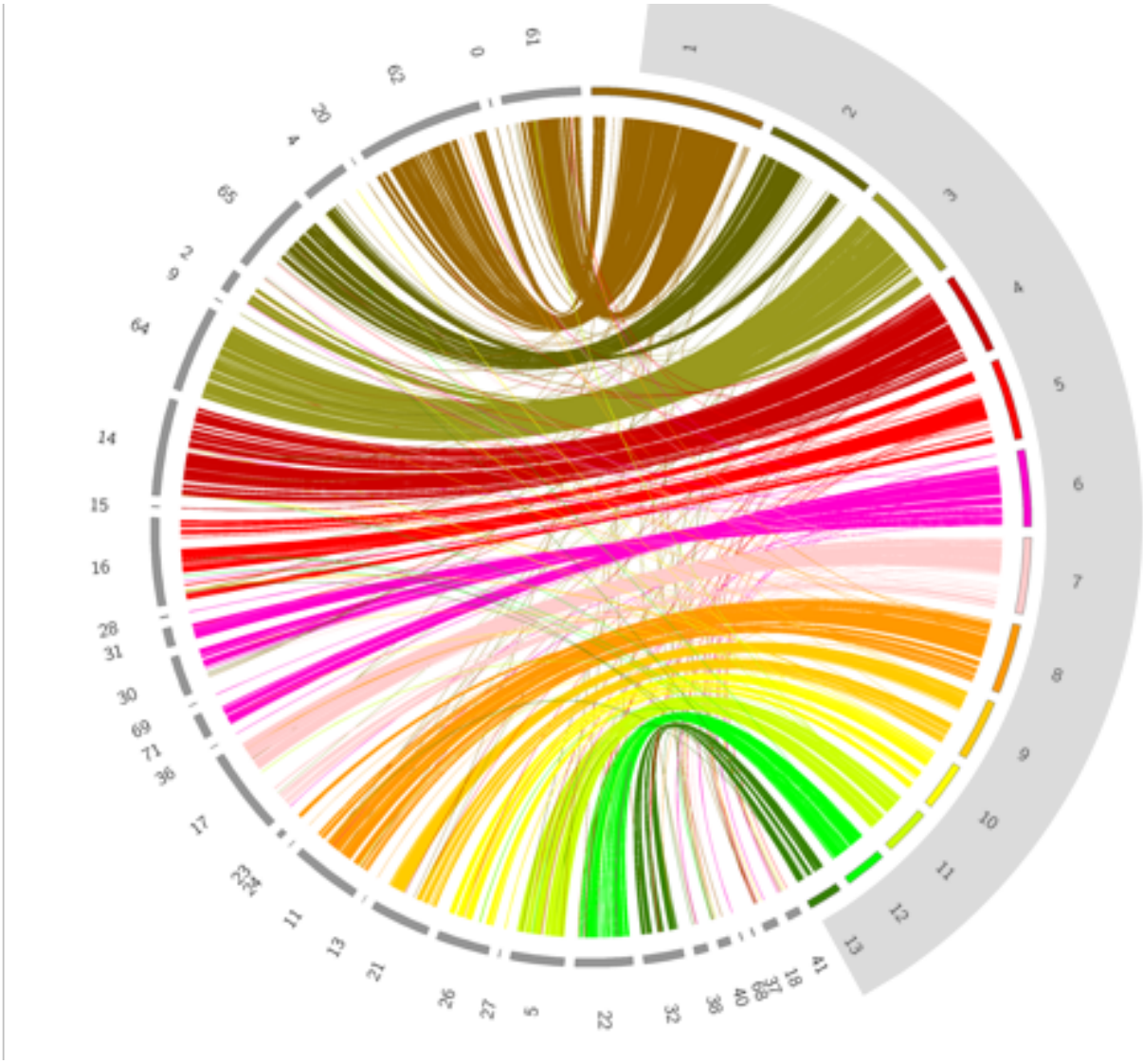


Figure S 9 - Circular synteny plot of the unitigs of Zb87 and the core chromosomes of Zt09. The length of the synteny blocks increases between Zb87 and Zt09 compared to the comparisons with Za17 and Zpa63. However, the Synteny is also interrupted by translocations between Zb87 and Zt09, although these are shorter than those in Za17.

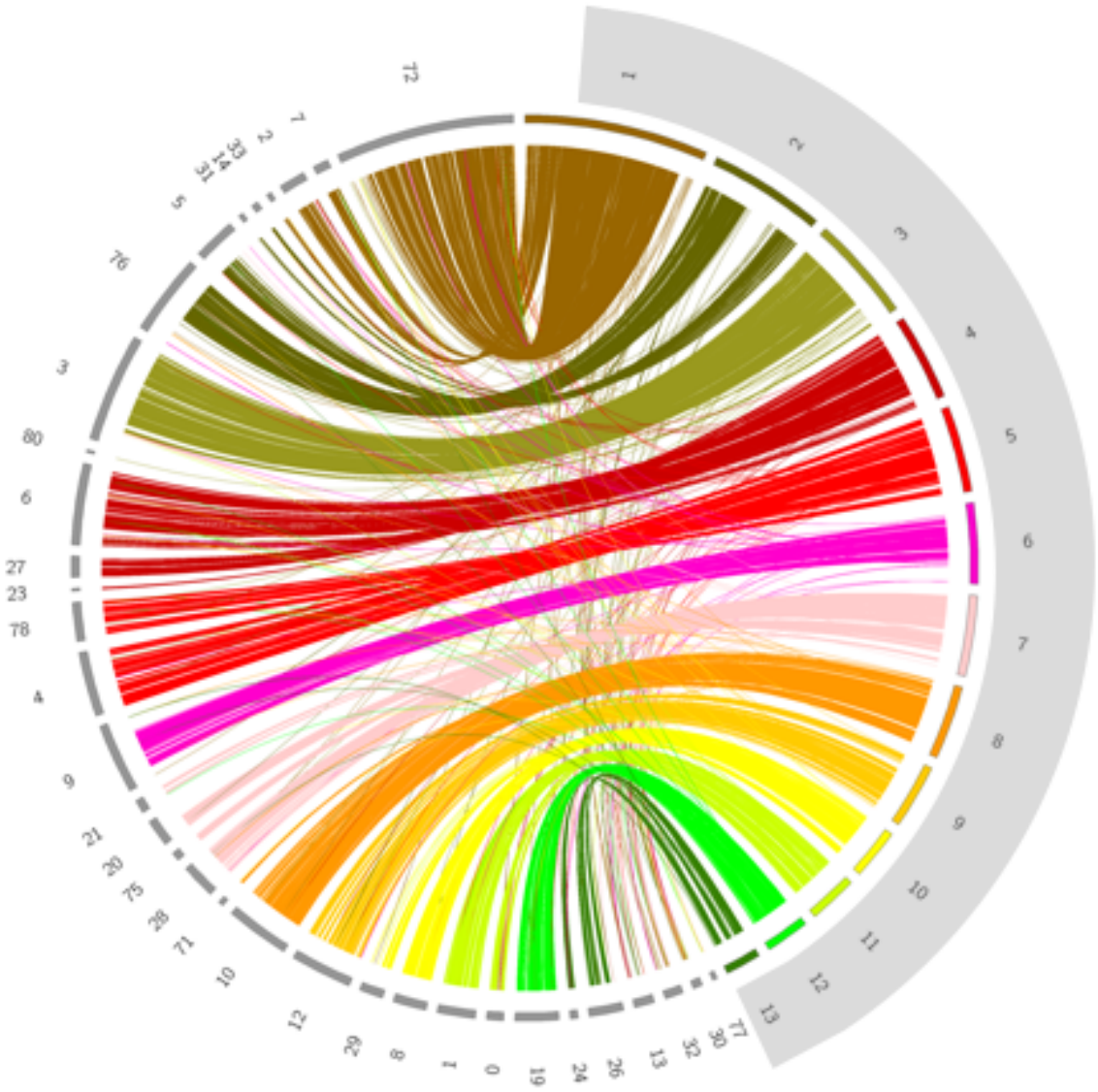


Figure S 10 - Circular synteny plot of the unitigs of Zp13 and the core chromosomes of Zt09. Like the comparison between Zb87 and Zt09, the comparison with Zp13 shows a high synteny with long synteny blocks interrupted by translocations of shorter sequence segments.

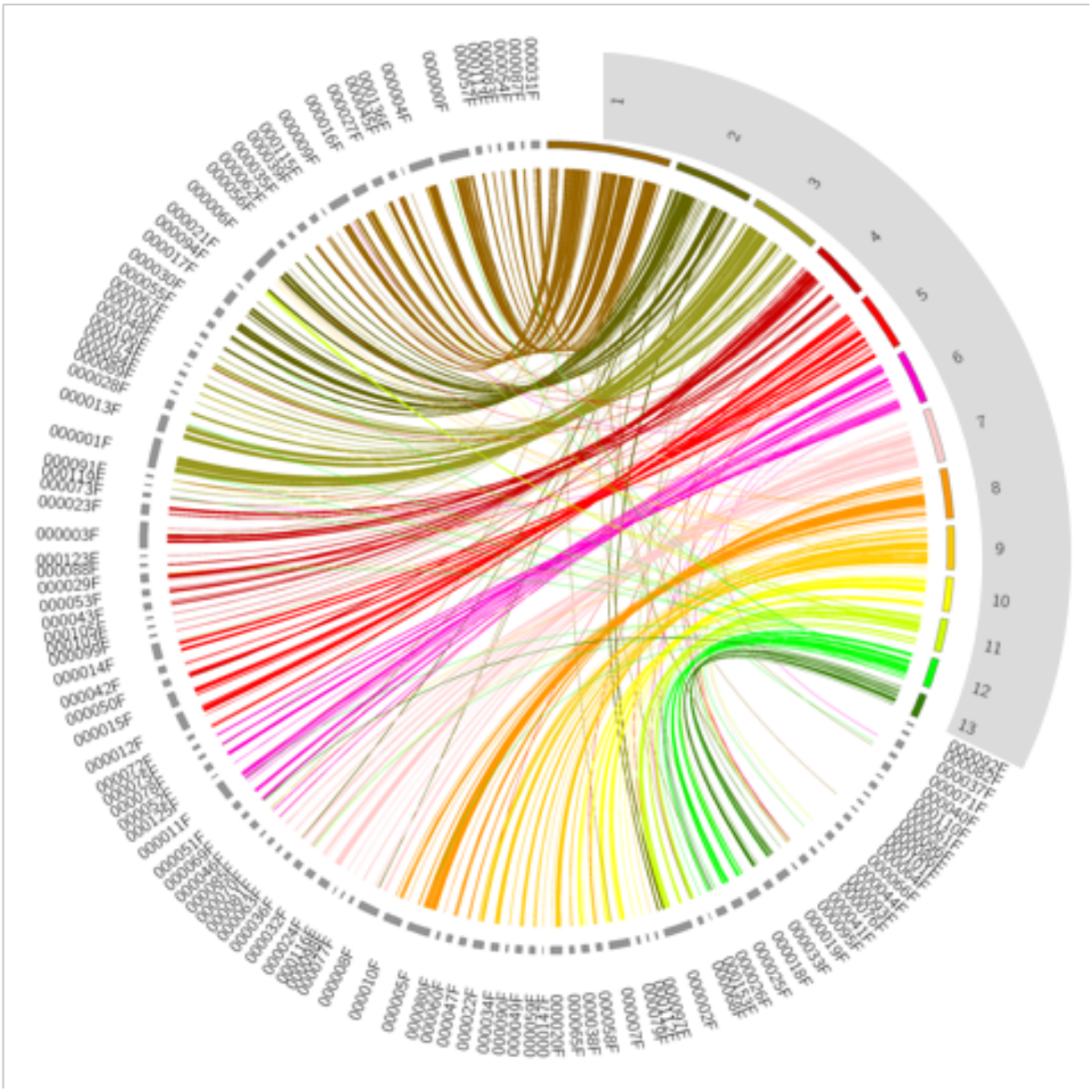


Figure S 11 - Circular synteny plot of the unitigs of Zpa63 and the core chromosomes of Zt09. Compared to the other synteny comparisons, the one between Zpa63 and Zt09 has the shortest synteny blocks. However, it seems likely that this is related to the greater fragmentation of Zpa63 genome assemblies.

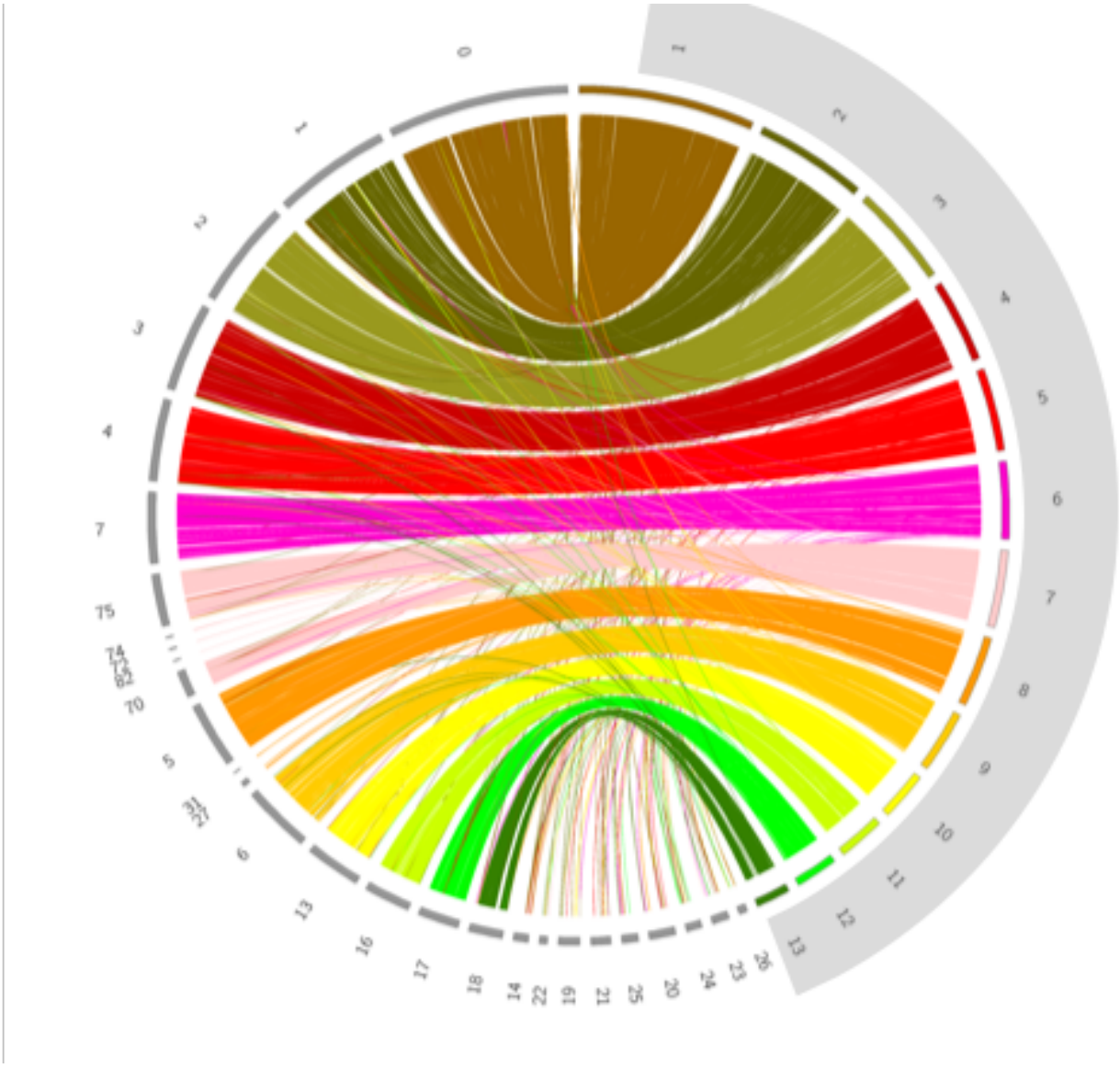


Figure S 12 - Circular synteny plot of the unitigs of Zt05 and the core chromosomes of Zt09. As expected, the intra-specific synteny comparisons (Zt05 against Zt09) show the highest level of synteny. However, here synteny blocks are interrupted by smaller rearrangements.

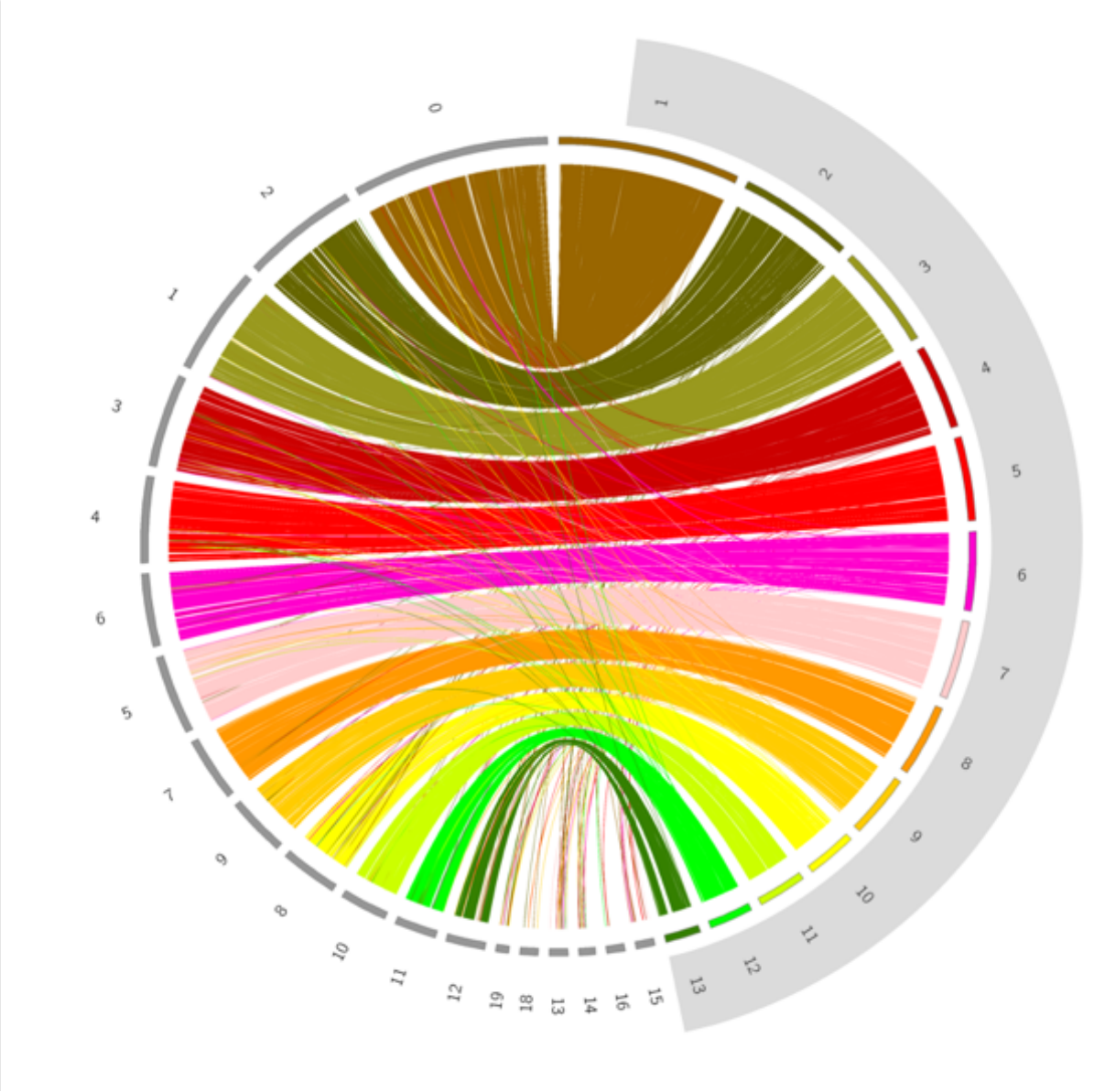


Figure S 13 - Circular synteny plot of the unitigs of Zt10 and the core chromosomes of Zt09. As shown by the comparison between Zt05 and Zt09, the comparisons at the intra-species level show the highest synteny. However, here synteny blocks are interrupted by smaller rearrangements.

8.6.3. Detailed circular synteny plots of the accessory chromosomes of Zt09 with the wild grass associated sister species

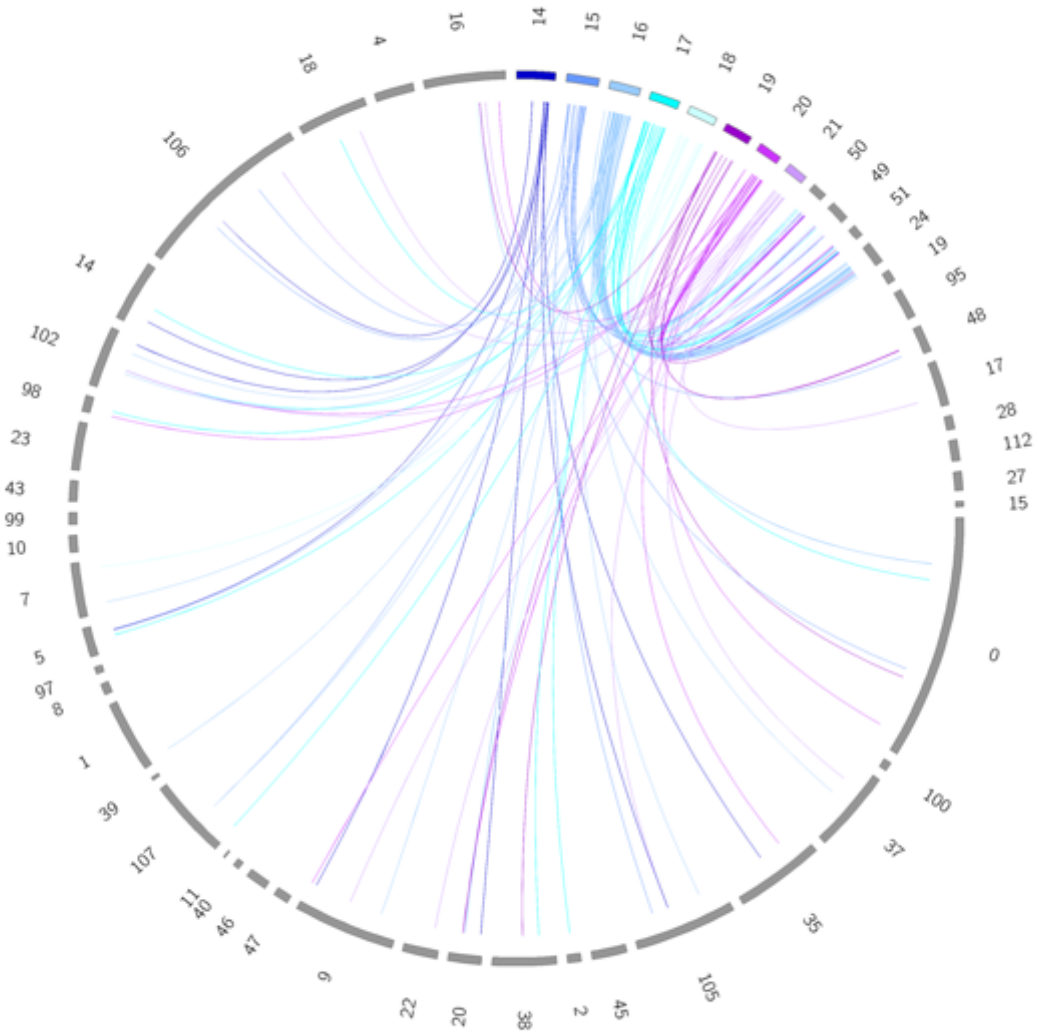


Figure S 14 - Circular synteny plot of the unitigs of Za17 and the accessory chromosomes of Zt09. The circular synteny plot between Za17 and the accessory chromosomes of Zt09 indicates that some of the *Z. tritici* accessory sequences were already present in the last common ancestor with *Z. ardabiliae*. It also shows that some of the sequences are still present in synteny blocks.

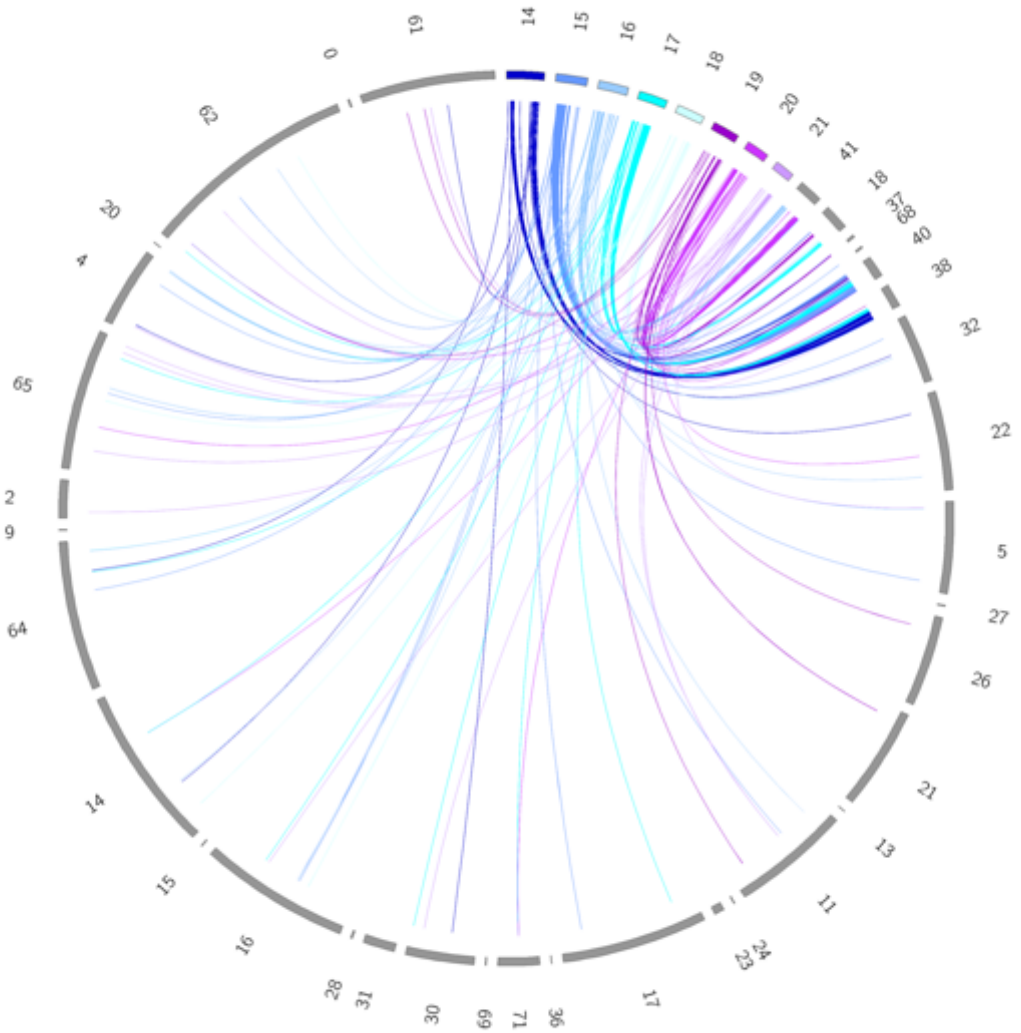


Figure S 15 - Circular synteny plot of the unitigs of Zb87 and the accessory chromosomes of Zt09. The pairwise comparison between the genome of Zb87 and the accessory chromosomes of Zt09, as well as the comparison with Za17, shows that some of the accessory sequences are inter-specific shared and in synteny. As already shown in the synteny analyses of the core genome, longer synteny blocks between Zb87 and Zt09 are also to be found in the accessory part of the genome than in the comparisons with Za17 or Zpa63.

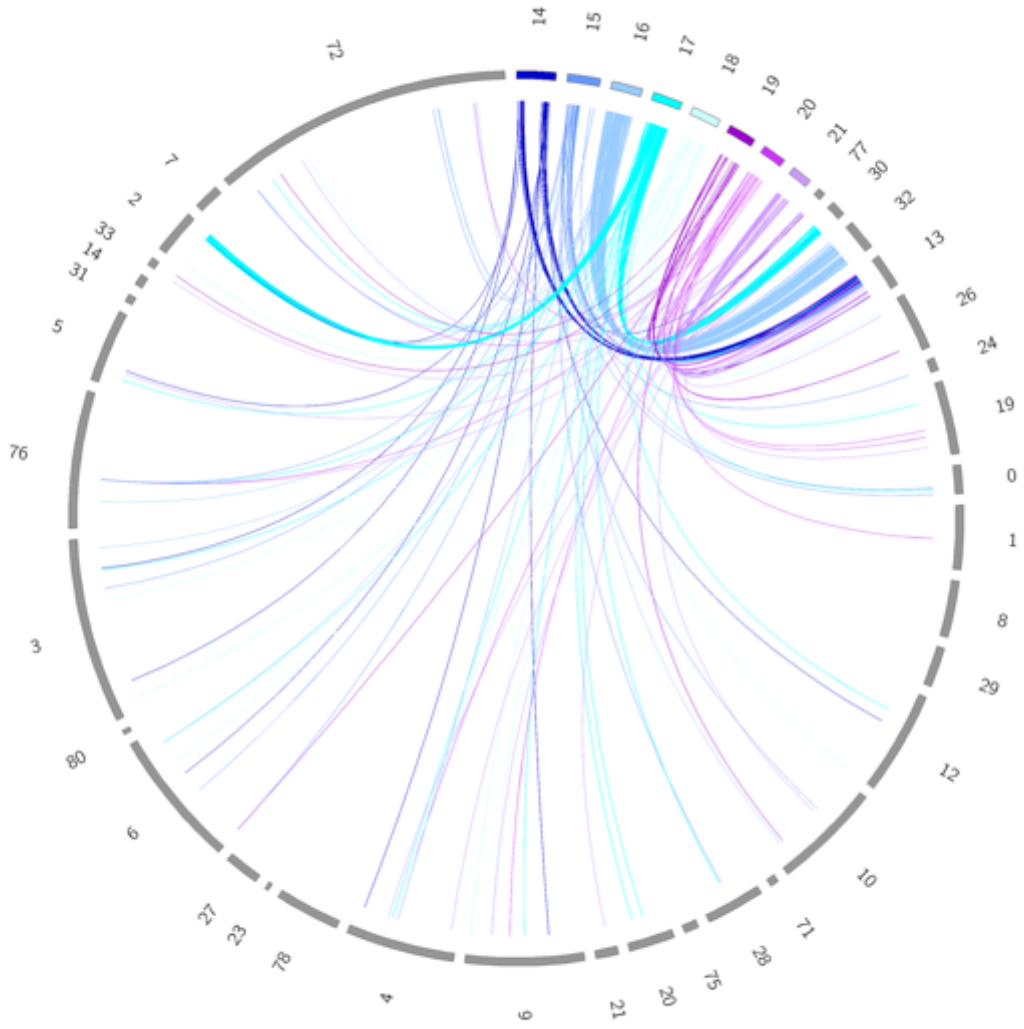


Figure S 16 - Circular synteny plot of the unitigs of Zp13 and the accessory chromosomes of Zt09. The accessory part of the genome of *Z. tritici* shows a high synteny with some unitigs of Zp13. This suggests that accessory chromosomes can also be shared as a syntenic unit between closely related *Zymoseptoria* species. Thus, the accessory chromosome 16 of Zt09 has an increased synteny with the unitig 32 of Zp13.

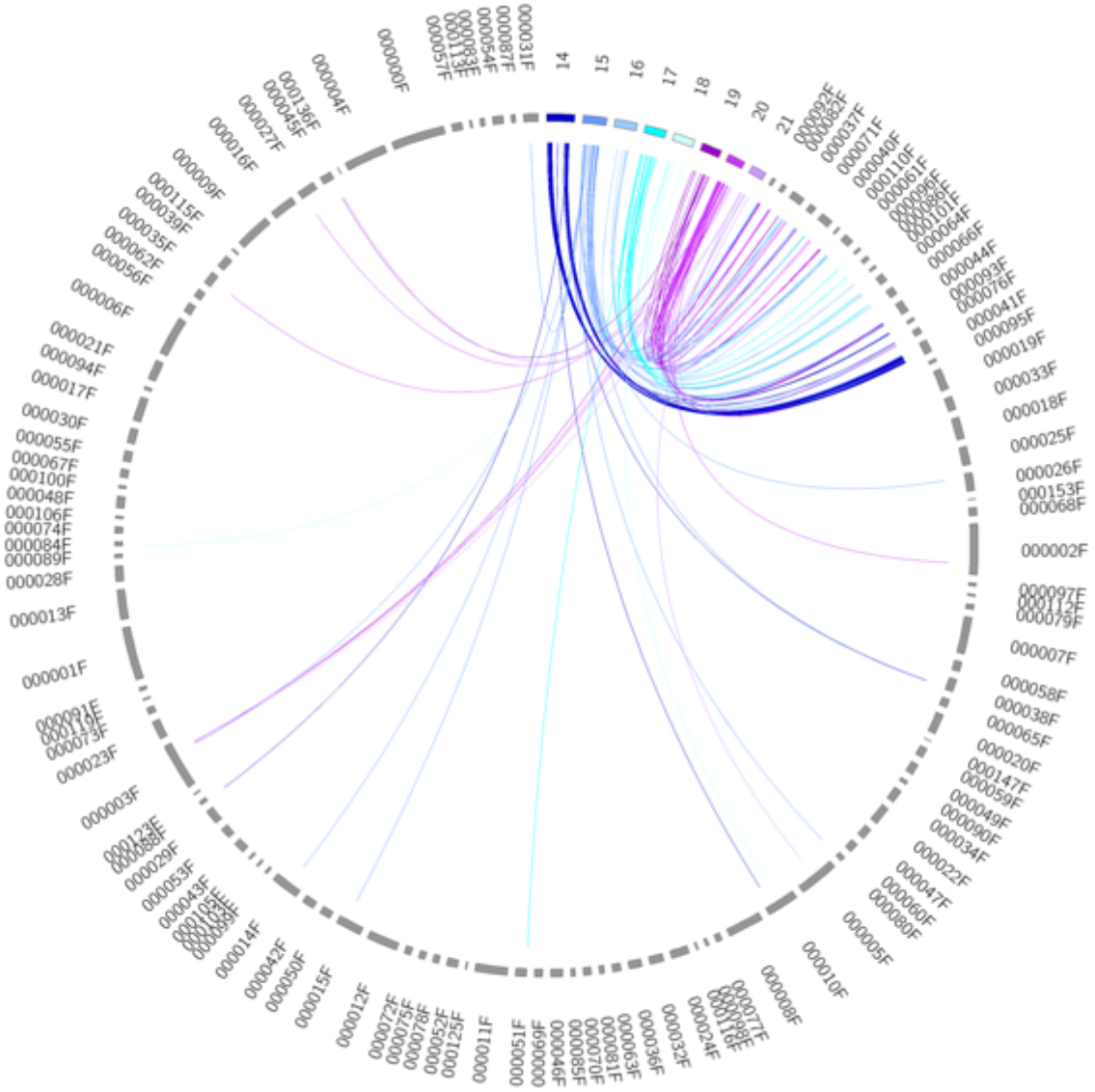


Figure S 17 - Circular synteny plot of the unitigs of Zpa63 and the accessory chromosomes of Zt09. Also in the comparison between Zt09 and the outgroup Zpa63, accessory sequences show synteny blocks in the closely related barley pathogen (Zpa63). The shorter length of the blocks could be due to the intense fragmentation of the assembly.

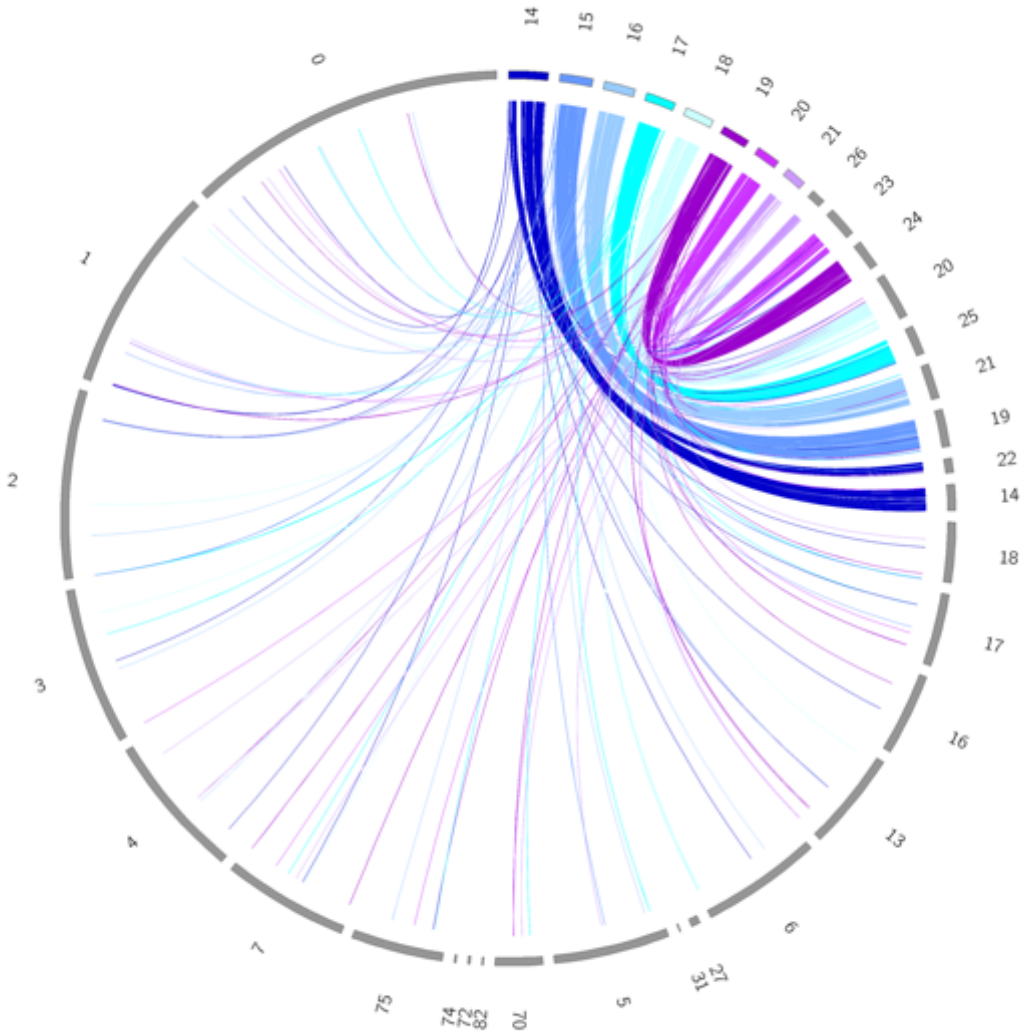


Figure S 18 - Circular synteny plot of the unitigs of Zt05 and the accessory chromosomes of Zt09. The ideogram indicates that all eight accessory chromosomes of Zt09 have homologs in Zt05.

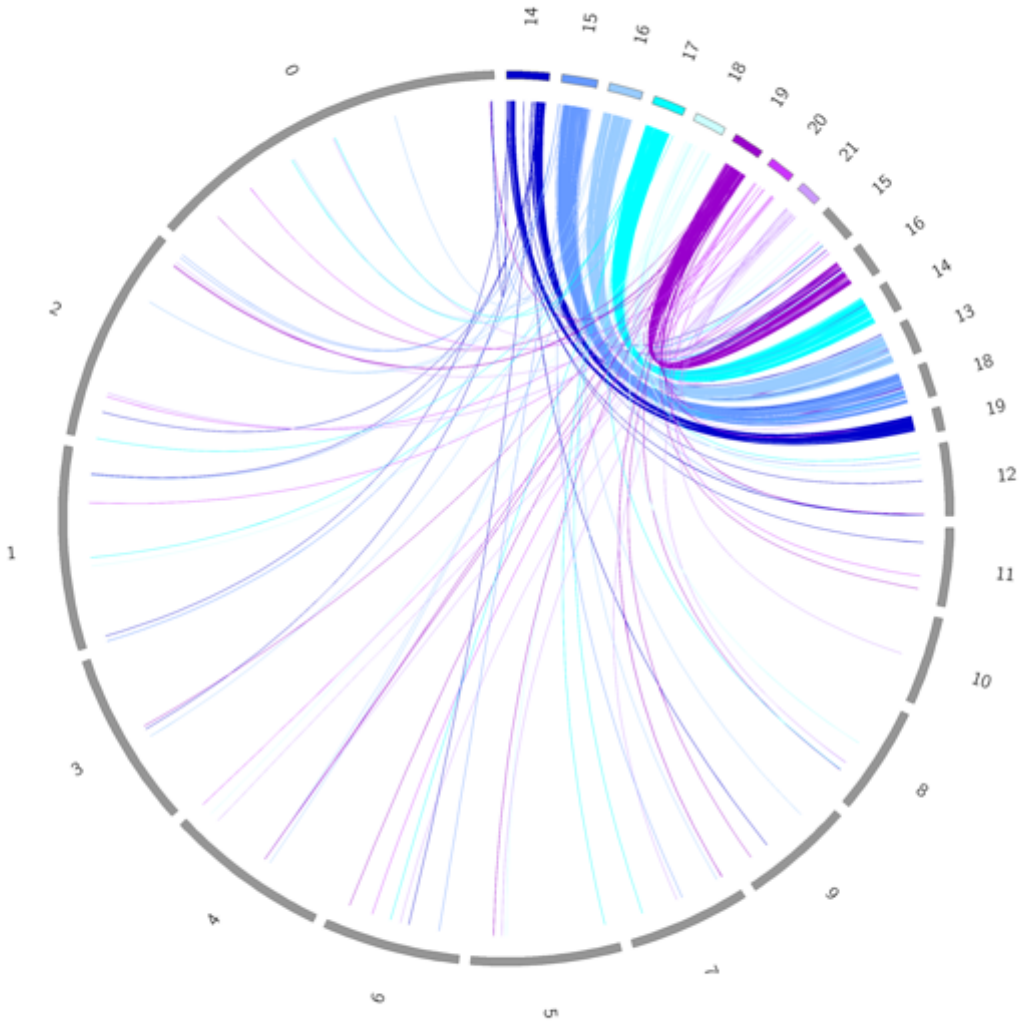


Figure S 19 - Circular synteny plot of the unitigs of Zt10 and the accessory chromosomes of Zt09. The figure indicates the absence of a homolog of the accessory chromosome 18 of Zt09 in the sister strain Zt10.

8.7. A detailed description of the detection of structural variants

The *de novo* assembled and repeat masked genomes of Za17, Zb87, Zp13, Zpa63, Zt05, Zt09, and Zt10 were indexed for read alignment using BWA "*index"* (Li and Durbin 2009) and for re-alignment using Novoalign "*index*" (www.novocraft.com). Alignment of short read sequencing data of 16 *Z. ardabiliae*, 8 *Z. brevis*, 26 *Z. pseudotritici,* and 12 *Z. tritici* isolates was performed using BWA "*aln*" (Li and Durbin 2009) with 16 processors ("*-t*") for forward and reverse reads separately. The alignment results were merged using BWA "*sampe*" (Li and Durbin 2009) and converted from SAM to BAM format using Samtools "*view"* with the parameter *"-Sb*" (H. Li et al. 2009). Discordant reads were extracted using Samtools "*view*" using the parameter setting "*-uF 1294"* and converted into a sequencing read format (FASTQ) using bamToFastq (Lindenbaum 2015). Extracted reads were re-mapped on the respective reference genome using Novoalign with sensitive settings. For re-alignment the median fragment library size was set to 500 with a standard deviation of 50 ("*-i*"), the repeat report strategy ("*-r*") was set to exhaustive mode with an upper limit of 1100 alignments allow moreover, the maximal alignment score threshold was set to 300 ("*-t*") which basically means that at least 10 mismatches were allowed.

8.8. Alignment statistics of the pairwise genome alignments

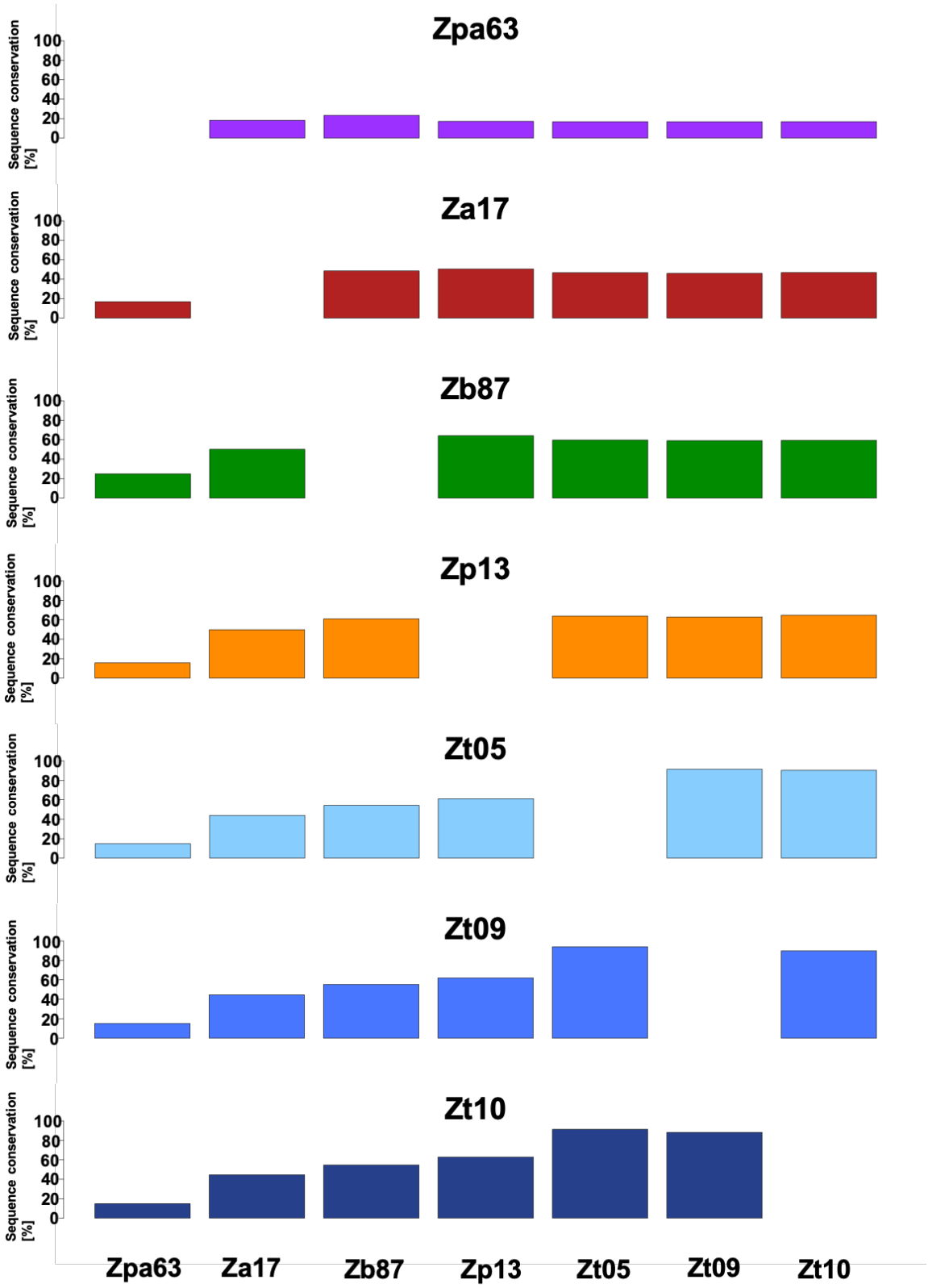


Figure S 20 - Total length of conserved regions. The y-axis shows the proportion of the genome that is covered by the alignment (genome without TEs) in percent. The graphs and their x-axes are ordered according to the phylogenetic relationships with an increasing distance from top to bottom (graphs) and from left to right (x-axes). Empty entries indicate the reference strain.

8.9. Correlation between pairwise synteny breaks and the positions of repetitive sequences in the *Zymoseptoria* species complex.

Table S 3 - Summary of the correlation studies between pairwise synteny breaks and repetitive genome content in a reference. sh = significant correlation. ns = no significant correlation. The table shows the result of the pairwise comparisons of the correlation studies between synteny breakpoints and the position of repetitive sequences. The results show that the majority of the comparisons were significant. However, the comparisons with Zpa63 as a query showed no significant correlation.

	Reference						
	Zpa63	**Za17**	**Zb87**	**Zp13**	**Zt05**	**Zt09**	**Zt10**
Zpa63		ns	ns	sh	ns	ns	ns
Za17	ns		sh	sh	sh	sh	sh
Zb87	ns	sh		sh	sh	sh	sh
Zp13	sh	sh	sh		sh	sh	sh
Zt05	sh	sh	sh	sh		sh	sh
Zt09	sh	sh	sh	sh	sh		sh
Zt10	sh	sh	sh	sh	sh	sh	

8.10. Alignment statistics of the multiple genome alignments

8.10.1. Alignment statistics of the intra-specific multiple genome alignments

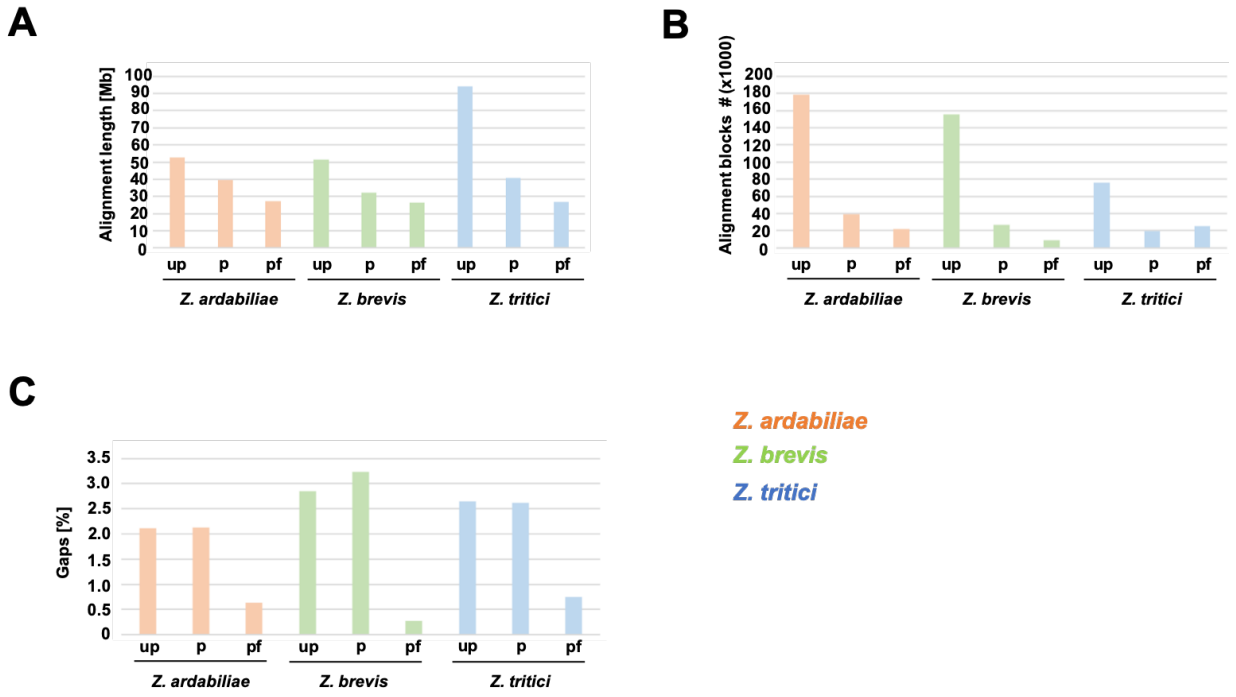


Figure S 21 - Alignment statistics of the intra-species MGAs of *Z. ardabiliae*, *Z. tritici,* and *Z. brevis*. The total block length (A), the total number of alignment blocks (B), as well as the proportion of gaps (C), have been computed for the raw (up), projected (p) and filtered (pf) versions of the MGAs using a custom python script ("*QMS.py*").

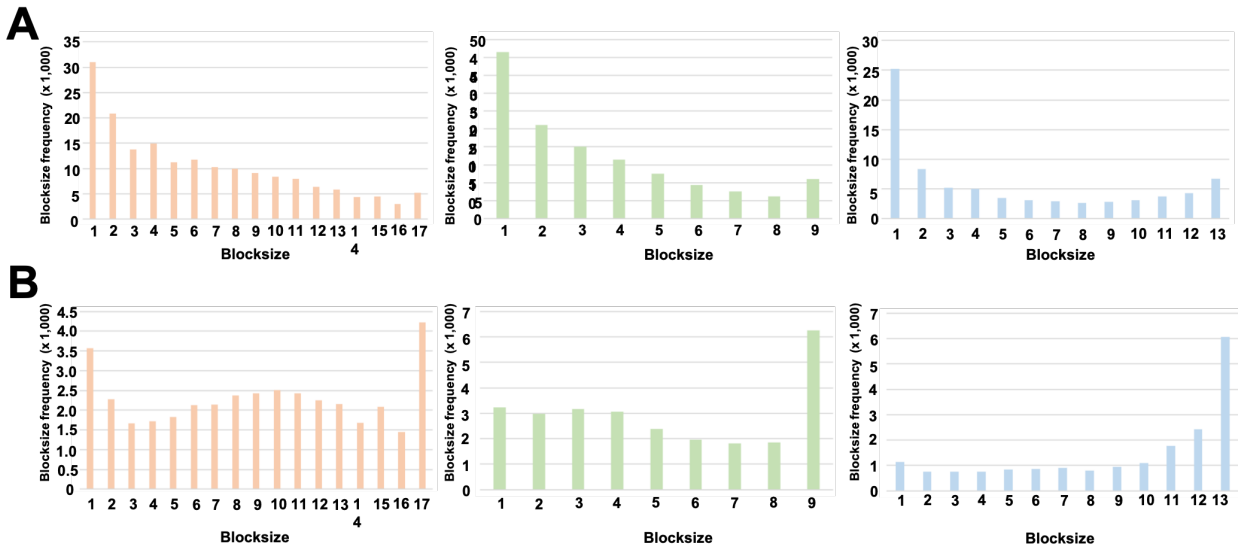


Figure S 22 - Blocksize frequencies in the MGAs of *Z. ardabiliae* (red), *Z. brevis* (green) and *Z. tritici* (blue). Frequencies were calculated for the raw (A) and projected (B) versions of the MGAs using a custom python script ("*QMS.py*"). Blocksize = number of sequences in alignment block.

8.10.2. Alignment statistics of the inter-specific multiple genome alignments

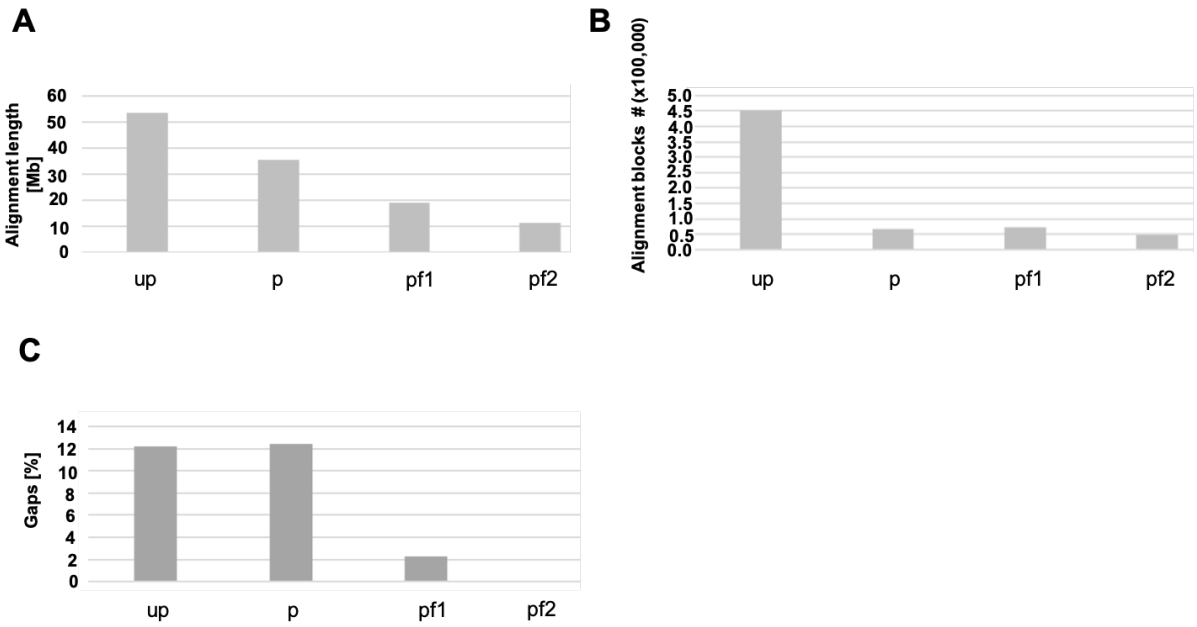


Figure S 23 - Alignment statistics of inter-specific multiple genome alignment. The figure shows the alignment length (A), the number of alignment blocks (B) and the percentage of gaps (C) for all steps of the alignment post-processing procedure. Up = unprojected, p = projected, pf = projected & filtered.

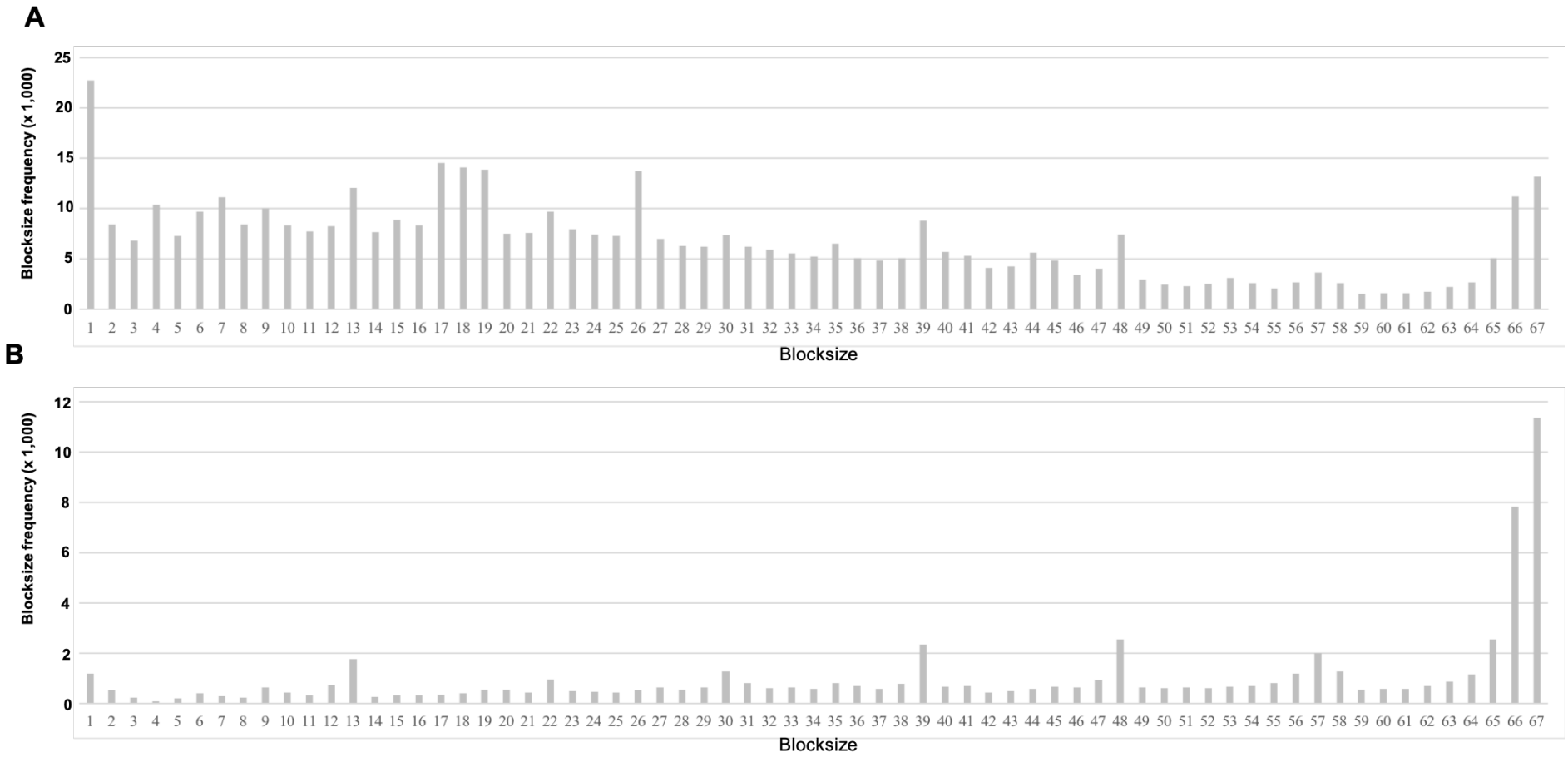


Figure S 24 - Alignment statistics of inter-specific multiple genome alignment. The figure illustrates the distribution of block sizes before (A) and after the projection (B) against a reference genome. Blocksize = number of sequences in alignment block.

8.11. Principle component analyses of *Z. ardabiliae, Z. brevis and Z. tritici*

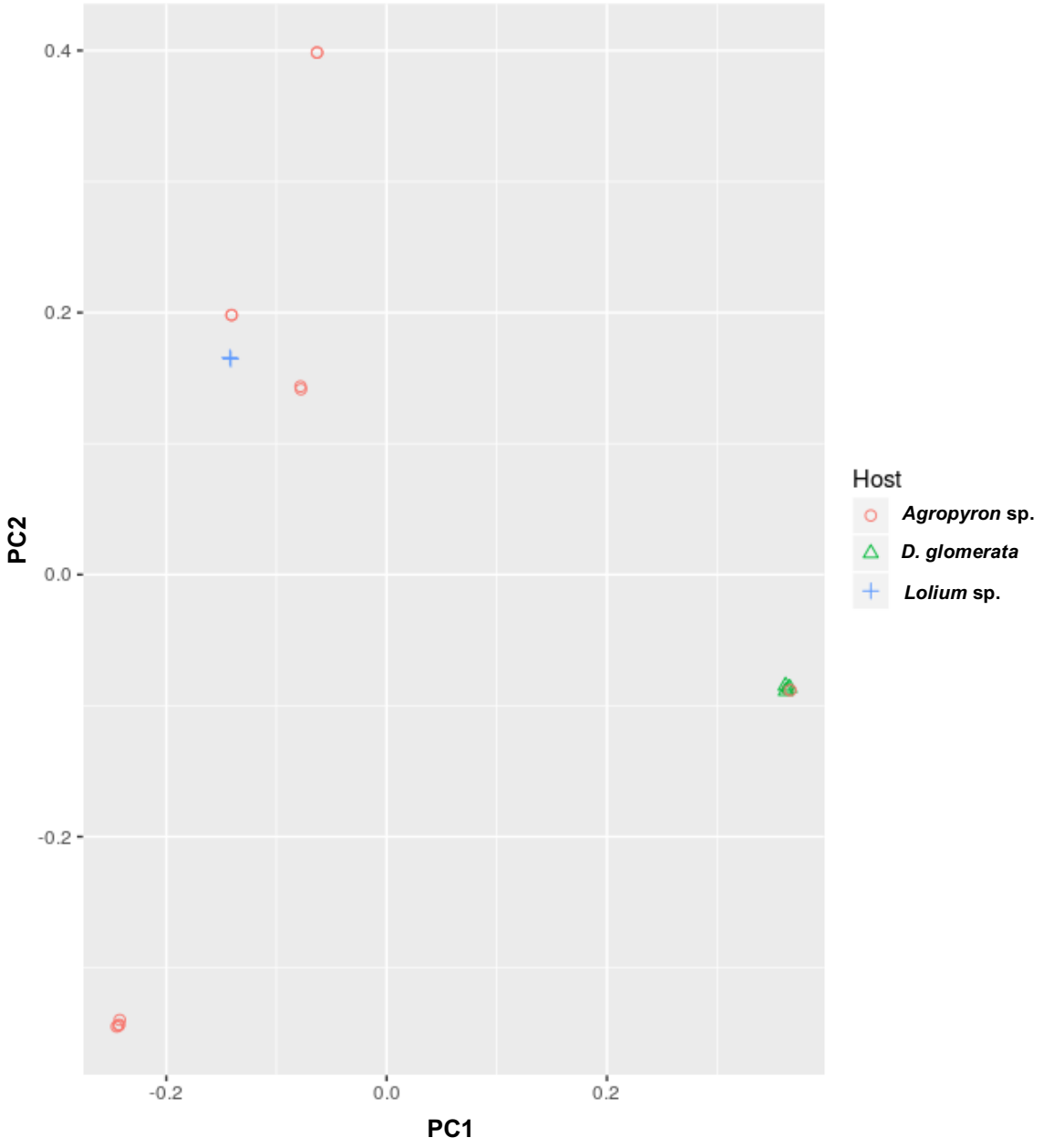


Figure S 25 - Principle component analysis of the *Z. ardabiliae* dataset. The analysis shows no grouping according to the original host. PC = Principal component.

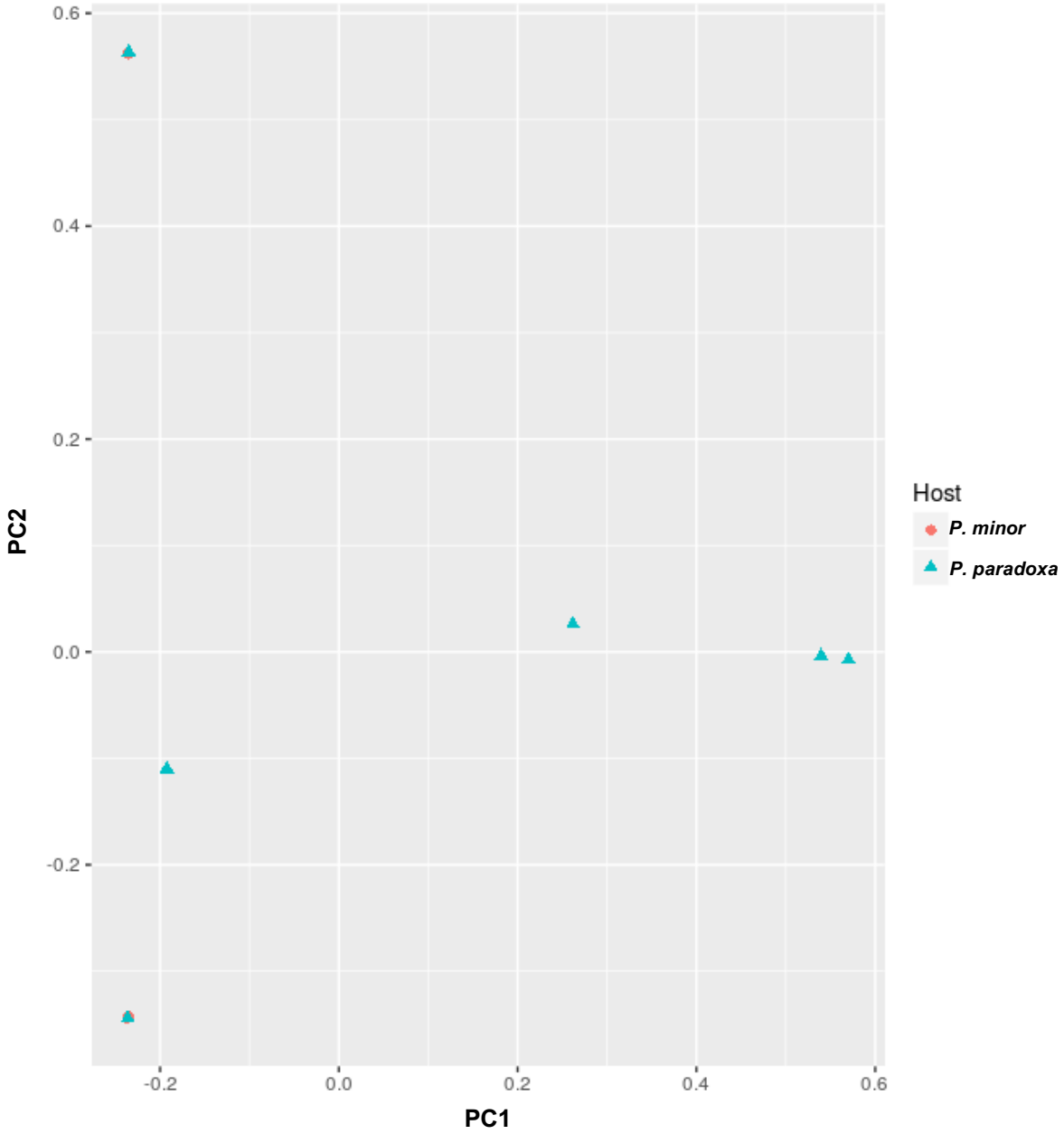


Figure S 26 - Principle component analysis of the *Z. brevis* dataset. The analysis shows no grouping according to the original host. PC = Principal component.

8.12. Analysis of the linkage disequilibrium in the *Zymoseptoria* species complex

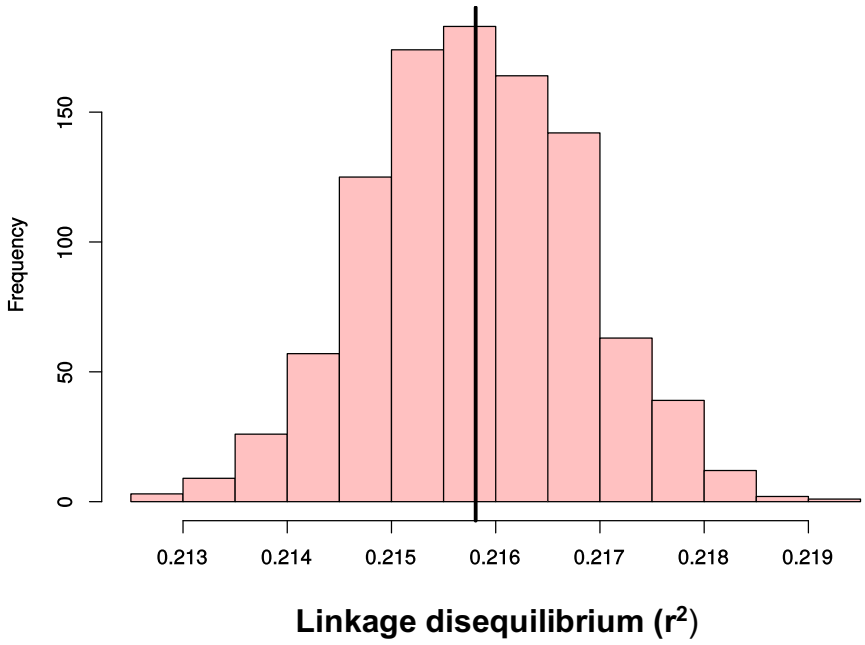


Figure S 27 - The histogram shown here presents the distribution of the mean values of the parameter r2 in *Z. ardabiliae.* They were calculated by each of the 1000 repetitions of the reduction of the data sets. It shows that there is a normal distribution between the values 0.213 and 0.219. The black vertical line marks the position of the representative example shown in the central part of this thesis in this distribution.

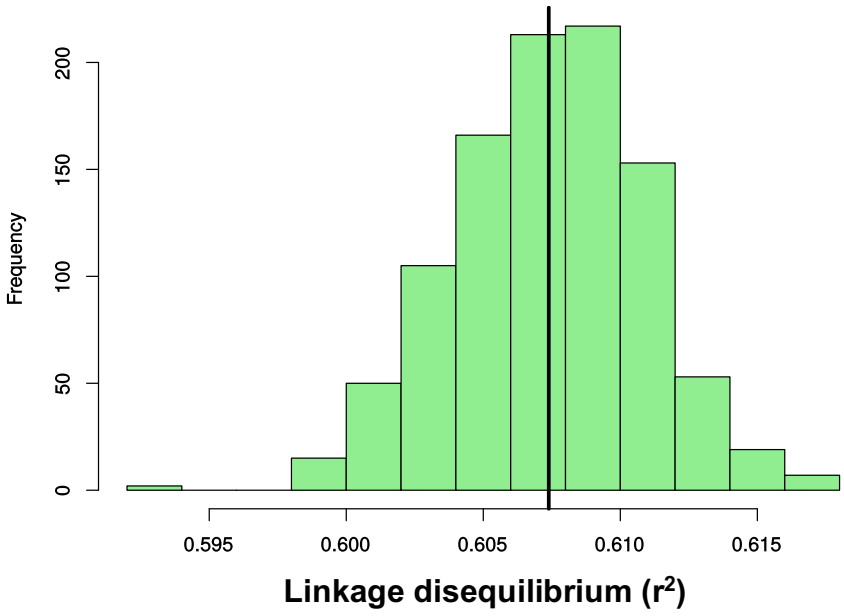


Figure S 28 - The histogram shown here presents the distribution of the mean values of the parameter r2 in *Z. brevis.* They were calculated by each of the 1000 repetitions of the reduction of the data sets. It shows that

there is a normal distribution between the values 0.595 and 0.615. The black vertical line marks the position of the representative example shown in the central part of this thesis in this distribution.

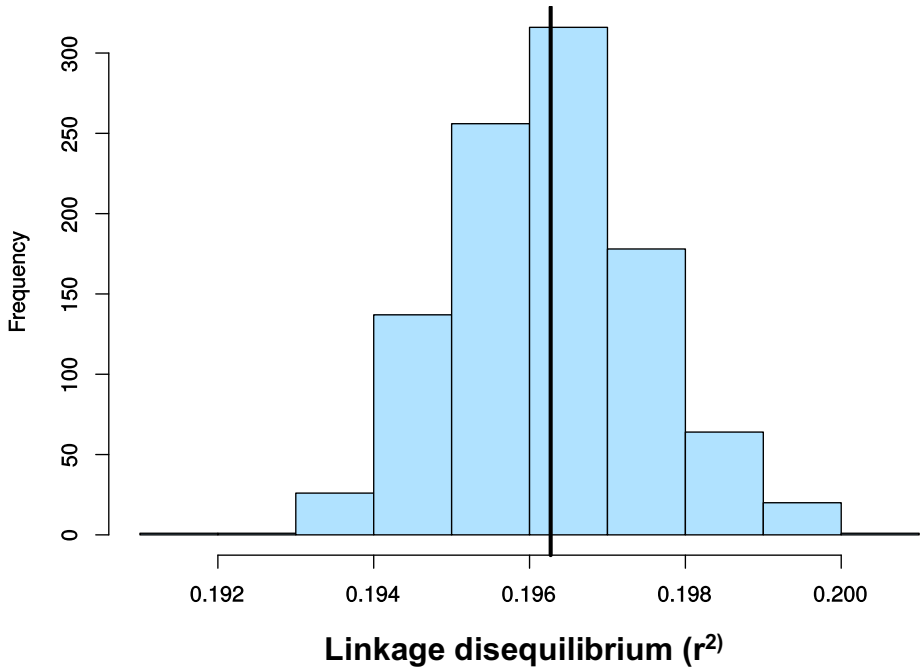


Figure S 29 - The histogram shown here presents the distribution of the mean values of the parameter r2 in *Z. tritici*. They were calculated by each of the 1000 repetitions of the reduction of the data sets. It shows that there is a normal distribution between the values 0.595 and 0.615. The black vertical line marks the position of the representative example shown in the central part of this thesis in this distribution.

8.13. Demographic analyses of the *Zymoseptoria* species complex using MSMC

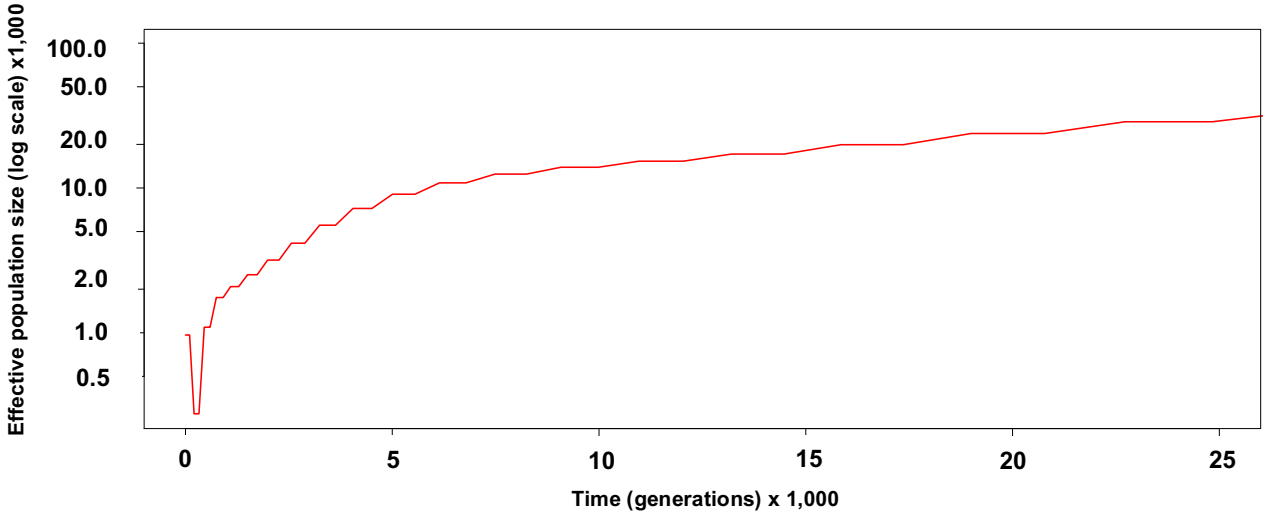


Figure S 30 - Demographic history calculated on the isolates Za91, 97 and 98 (Pop1). The graph shows a recent bottleneck effect in *Z. ardabiliae*.

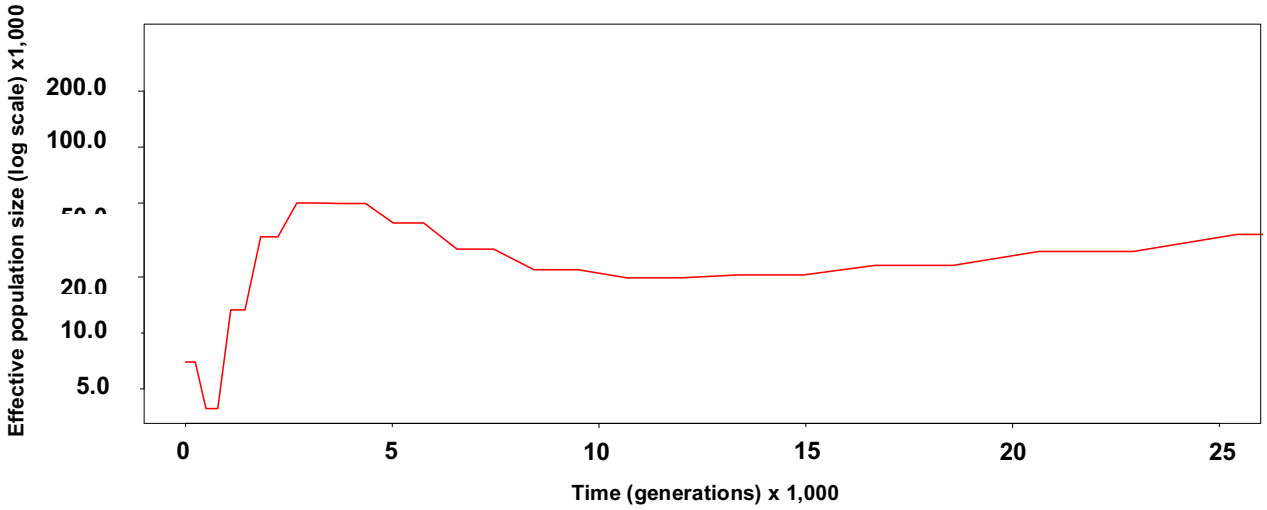


Figure S 31 - Demographic history calculated on the isolates Za17, 20, 95 and 96 (Pop2). The graph shows a recent bottleneck effect in *Z. ardabiliae*.

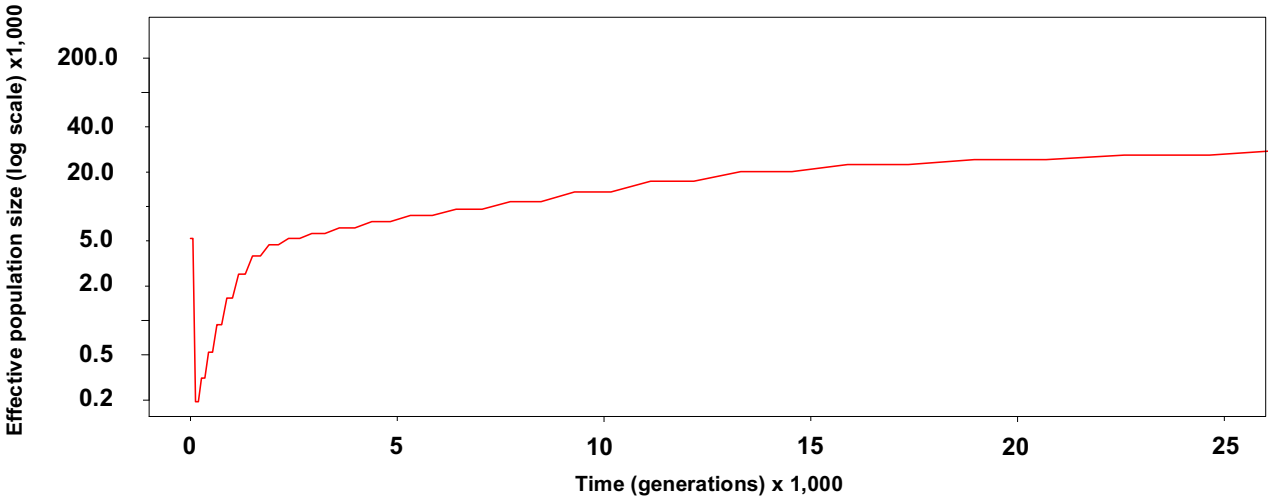


Figure S 32 - Demographic history calculated on the isolates Za90, 92 and 93 (Pop3). The graph shows a recent bottleneck effect in *Z. ardabiliae*.

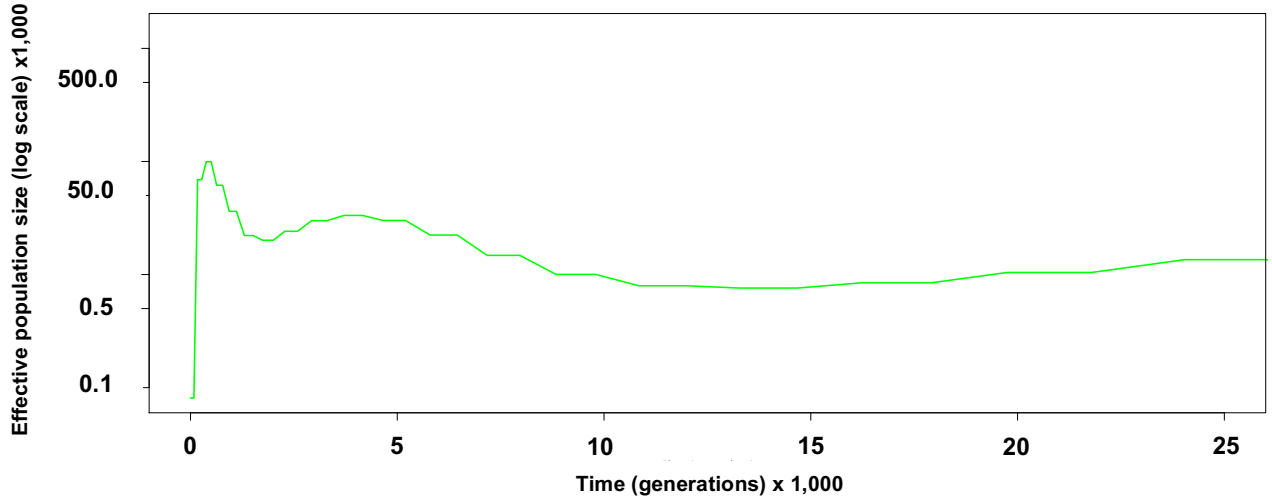


Figure S 33 - Demographic history calculated on the isolates Zb85, 86, 161 and 163. The graph shows a recent population size expansion in *Z. brevis*.

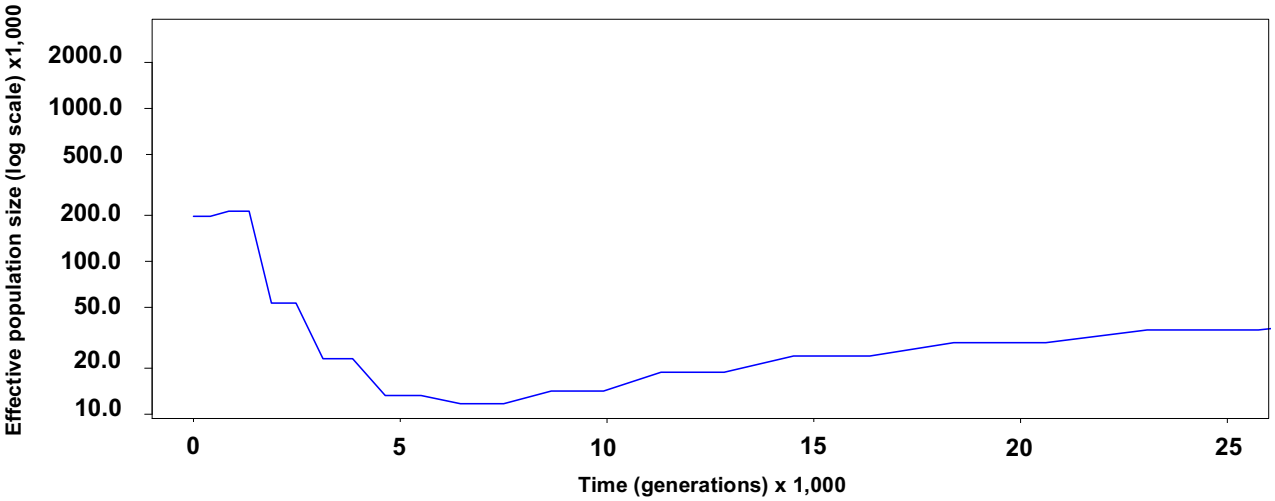


Figure S 34 - Demographic history calculated on the isolates Zt05, 09, 151 and 155 (only Europe). The graph shows a substantial population size expansion in *Z. tritici*.

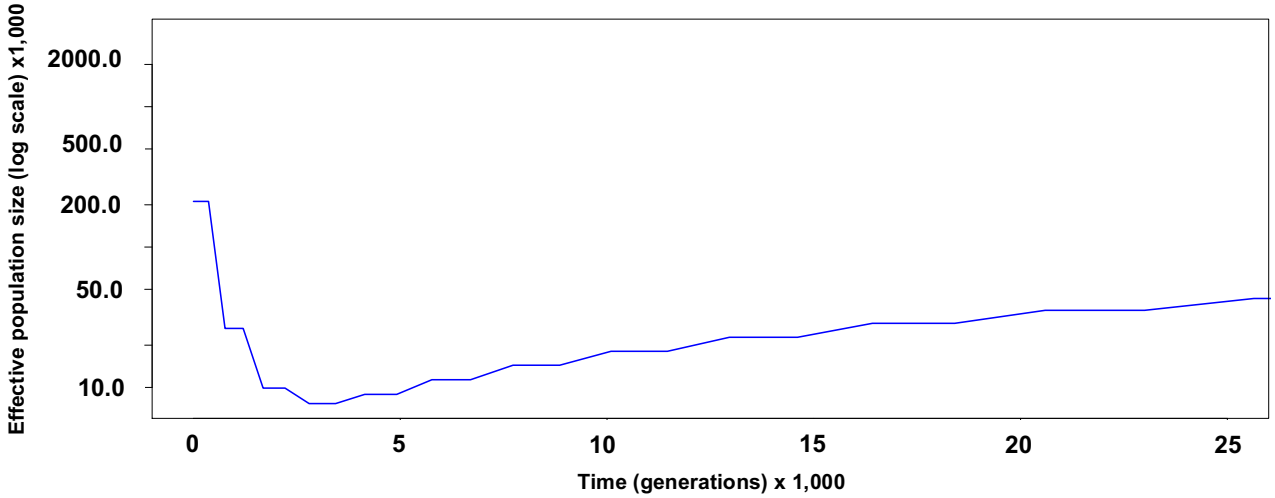


Figure S 35- Demographic history calculated on the isolates Zt02, 04, 05 and 07 (only Denmark). The graph shows a substantial population size expansion in *Z. tritici*.

8.14. Signatures of positive diversifying seletction in *chr_7_00125*

Table S 4 - Sites under positive diversifying selection in *chr _7_00125* as detected by the BEB methods implemented in M2a and M8. SE = standard error, aa = amino acid, BEB = Bayes empirical Bayes

Position [aa]	Substitution	Model M2a (BEB)		Model M8 (BEB)	
		Mean ω +/- SE	[Pr(w>1)]	Mean ω +/- SE	[Pr(w>1)]
109	Q → K	3.816 +- 0.995	0.986*	2.908 +- 0.960	0.995**
236	R → Q	3.698 +- 1.140	0.940	2.868 +- 1.010	0.973*
459	S → G	2.561 +- 1.594	0.583	2.284 +- 1.347	0.749
668	S	3.785 +- 1.038	0.973*	2.900 +- 0.972	0.990*
1,220	P → Q	3.054 +- 1.453	0.740	2.725 +- 1.110	0.918
1,285	S → T	3.339 +- 1.347	0.831	2.834 +- 1.027	0.962*
1,360	R	3.725 +- 1.153	0.958*	2.805 +- 1.124	0.941
1,363	L	2.875 +- 1.909	0.713	2.098 +- 1.639	0.649
1,403	S	3.193 +- 1.460	0.777	2.668 +- 1.176	0.890
1,456	L	2.286 +- 2.059	0.559	1.717 +- 1.696	0.518
1,457	K → N	3.284 +- 1.368	0.814	2.825 +- 1.033	0.958*
1,490	P → G	3.101 +- 1.528	0.744	2.551 +- 1.262	0.841
1,492	G → P	3.129 +- 1.519	0.753	2.565 +- 1.255	0.846
1,500	R → E	----------	--------	2.088 +- 1.375	0.682

Table S 5 - Protein domains in *chr_7_00125* as predicted by the Coach server. The C-score determines the confidence as values between 0 and 1. Higher values indicate higher confidence. aa = amino acid.

Domain ID	Position in the sequence [aa]	C-score	Cluster size	Ligand Name	Positions of the putative binding residues [aa]
Transmembrane region	214 - 236	-----	-----	-----	-----
Transmembrane region	256 - 278	-----	-----	-----	-----
ABC-Membrane	353 - 529	0.09	5	GSH[a]	493; 496; 497; 500; 519; 520; 522; 523; 524
Coiled-coil region	598 - 629	-----	-----	-----	-----
AAA-ATPase	655 - 828	0.94	229	ADP[b]	666 - 672
Low complexity region	855 - 879	-----	-----	-----	-----
ABC-Membrane	917 – 1,183	0.05	3	0JZ[c]	932; 935; 936; 939; 1,150; 1,173; 1,177; 1,180
AAA-ATPase	1,256 – 1,451	0.89	220	ADP[b]	1,267-1,273

a) GSH = Glutathione

b) ADP = Adenosine diphosphate

c) 0JZ = (4R,5R,11R,12R,18R,19S)-4,11,18-TRIS(1-METHYLETHYL)-6,13,20-TRISELENA-3,10,17,22,23,24-HEXAAZATETRACYCLO[17.2.1.1~5,8~.1~12,15~]TETRACOSA-1(21),7,14-TRIENE-2,9,16-TRIONE

8.15. About the attached USB-Stick

The USB-drive attached to this thesis contains annotated versions of all scripts and parameter files mentioned in this dissertation. It also contains a folder with the following supporting results:

Supplementary material for 4.2:

→ all *de novo* genome assemblies used

Supplementary material for 4.2.3 and 8.9:

→ pairwise correlation studies between the position of synteny breaks and the position of repetitive sequences for all possible pairwise combinations of the *Zymoseptoria* reference assemblies (Zpa63, Za17, Zb87, Zp13, Zt05, Zt09, and Zt10).

Supplementary material for 4.3.1:

→ chromosome/unitig-wise distribution of Tajima's Pi, Watterson theta and Tajima's D.

Supplementary material for 4.4**:**

→ Output files of the Codeml screen for signatures of positive diversifying selection for each gene and model comparisons.

→ Final lists of genes with signatures of positive diversifying selection before and after FDR correction for the model comparisons M1a/M2a and M7/M8 and those shared between the models.

→ Final lists of genes with signatures of positive directional selection.

→ Final lists of effector candidates plus separate lists of candidates with signatures of positive diversifying selection

→ Final lists of species-specific genes including a total dataset with all genes and sub-datasets that contain only species-specific effector candidates or genes with signatures of positive diversifying selection.

Danksagung

Zu allererst möchte ich mich bei Eva Stukenbrock für die Möglichkeit bedanken meine Promotion als Teil ihrer Arbeitsgruppe durchführen zu.können. Vielen Dank für dein Vertrauen und deine Unterstützung, welche mich nach dem Abschluss meiner Masterarbeit bei dir in Marburg dazu ermutigt haben, den Schritt zu wagen, dir für die Promotion nach Kiel zu folgen. Ich bin dir sehr dankbar dafür, dass du mir die Chance gegeben hast meine Fähigkeiten in der Bioinformatik und der IT-Administration weiter auszubauen und zu verfeinern. Doch vor allem dein Rat, deine Unterstützung und die vielen hilfreichen Diskussionen waren maßgeblich für die Fertigstellung dieser Dissertation.

Des Weiteren möchte ich mich bei den Mitgliedern meines *thesis advisory committee,* Julien Yann Dutheil und Bernhard Haubold, für die Unterstützung, sowie die vielen Anregungen und Tipps bedanken.

Darüber hinaus möchte ich mich bei Herrn Prof. Dr. Leippe, Frau Prof. Dr Dagan und Herr Prof. Dr. Haubold dafür bedanken, dass sie sich dazu bereit erklärt haben Teil meines Prüfungskomittees zu sein.

Bei Daniel Croll möchte ich mich für die Möglichkeit bedanken, den Umgang mit dem Programm SMRTAssembly zu *de novo* Erstellung von Genomen aus PacBio-Daten als Gast seiner Arbeitgruppe zu erlernen.

Besonders bedanken möchte ich mich bei Alice, Lizel und Julian, die mir nicht nur in der Bioinformatik immer mit Rat und Tat zur Seite standen. Ich möchte mich darüber hinaus auch sehr bei Mareike, Janine, Heike, Alina und Michael bedanken, die mich stets bei Fragen zur Laborarbeit unterstützt und mir somit dabei geholfen haben den Bezug dazu nicht völlig zu verlieren.

Bei dir Andrea, möchte ich mich sehr für die große Hilfe in administrativen Belangen und die tollen Gespräche bedanken, die immer mal zu der ein oder anderen „Überraschungs-Muffin-Attacke“ geführt haben.

Auch bedanken möchte ich mich im Allgemeinen bei allen Mitgliedern der Arbeitsgruppe für den Rückhalt und das freundschaftliche kollegiale Miteinander, welche maßgeblich zu der angenehmen Arbeitsatmosphäre beigetragen haben.

Besonders möchte ich mich bei Sonja bedanken. Du hast mir durch deine bedingungslose Unterstützung und deine immense Geduld mit meinen Stimmungsschwankungen und Krisen die Kraft dafür gegeben aus der sehr beschwerlichen Zeit des „Zusammenschreibens“ nicht als komplettes Wrack hervorzugehen. Ganz im Gegenteil, du hast mir den Halt und die Motivation gegeben immer wieder aufzustehen und voranzugehen und nicht aufzugeben.

In diesem Zusammenhang möchte ich mich noch besonders bei meinen Eltern bedanken, die es mir durch ihre Unterstützung und ihren liebevollen Rat ermöglicht haben auch schwere Zeiten im „weit“ entfernten Kiel zu überstehen. Diese bedingungslose Unterstützung, die von euch und dem Rest der Familie gekommen ist, haben diese Dissertation erst möglich gemacht.

Ich möchte mich außerdem bei Tobias, Mario, Jan, Ellen, Julian, Carina und Franzi dafür bedanken, dass ihr mir immer wieder die Möglichkeit gegeben habt die Arbeit auch einmal hinter mir zu lassen und einfach ein schöne Zeit zu haben. Drüber hinaus bin ich euch sehr dankbar, dass ihr

auch Zeiten, in denen ich mich sehr auf meine Arbeit konentrieren musste stets zu mir gehalten habt.

Affidavit

I hereby declare that

- this dissertation concerning content and design is the product of my own work under guidance of my supervisor Prof. Dr. Eva H. Stukenbrock. I used no other tools or sources but the cited ones.
- this dissertation has been conducted and prepared following the Rules of Good Scientific Practice of the German Research Foundation.
- this dissertation has not been submitted elsewhere partially or wholly as part of a doctoral degree to another examining body, and no other materials are published or submitted for publication than indicated in the thesis.
- I have not yet been deprived of an academic degree

Kiel, 02.05.19 Christoph Jakob Eschenbrenner